

Tesis doctoral

**Fuentes de variabilidad individual en el ajuste de
modelos de árbol: el caso de las conjunciones
ilusorias**

Juan Ignacio Durán Pacheco

Directores: Juan Botella Ausina y Manuel Suero Suñe

Universidad Autónoma de Madrid

Agradecimientos

Elaborar un trabajo de este tipo termina involucrando a más personas de las que uno espera al comienzo. Todos ellos hacen su aportación, en ocasiones incluso sin apenas tener conciencia de ello. Tanto es así, que es probable que olvide algunas en este momento. Entre las que puedo recordar ahora, quería agradecer su aportación personal o profesional a:

Juan y Manuel, por su profesionalidad, humildad y paciencia. Vuestro trabajo y dedicación han sido y son un estímulo constante. Del mismo modo, a José Miguel por sus consejos, su brillantez intelectual y su cercanía. Espero seguir aprendiendo de vosotros y haber adquirido muchas de vuestras virtudes para el futuro.

Orfelio y Nani. Gracias por vuestra generosidad y por brindarme un buen puñado de oportunidades para alejarme, acercándome a la montaña.

Al equipo de Mikel, que adoptó a un extraño con tanta hospitalidad. A Pili y a Carlos, mis primeros compañeros y consejeros de investigación. A ellos les han seguido mis vecinos de despacho Ana, Blanca y el correlativo David.

A mis nuevas compañeras, Cristina y Carolina y al CUCC en general, que han hecho de Alcalá un lugar más acogedor.

Lorena y Silvia, por ser tan auténticas y estar siempre disponibles para afrontar lo que sea que venga.

Por último a mi familia, que ha capeado todo cambio de humor imaginable y quién sabe cuántas cosas más.

Índice

Resumen	1
Abstract	3
 Presentación	 5
 1. Introducción.....	 9
1.1. Posibilidades no contempladas en el ajuste a partir de datos agregados: el supuesto de independencia e idéntica distribución (i.i.d.).....	9
1.2. Estrategias de tratamiento de datos previas a la estimación y ajuste. Ventajas y limitaciones	11
<i>1.2.1. Datos agrupados y estimación de un conjunto de parámetros general ...</i>	<i>11</i>
<i>1.2.2. Datos individuales y estimación de un conjunto de parámetros por participante</i>	<i>12</i>
<i>1.2.3. Estrategia mixta: grupos de participantes distintos</i>	<i>12</i>
1.3. Modelos propuestos para formalizar diferencias individuales.....	13
<i>1.3.1. Modelos para diferencias continuas: efectos aleatorios y rasgo latente .</i>	<i>13</i>
<i>1.3.2. Modelos para diferencias discretas: clase latente</i>	<i>14</i>
1.4. Atención y conjunciones ilusorias	14
<i>1.4.1. El paradigma de presentación rápida de series visuales</i>	<i>17</i>
1.5. Modelos multinomiales de procesamiento en árbol (MPT)	18
<i>1.5.1. Categorías de respuesta en el paradigma PRSV</i>	<i>20</i>
<i>1.5.2. Modelos de árbol para formalizar las posiciones sobre la existencia de las CI</i>	<i>21</i>
1.6. Hipótesis y plan general de estudios.....	23
<i>1.6.1. Estudio 1. Respecto al debate sobre las CI</i>	<i>23</i>
<i>1.6.2. Estudio 2. Respecto al rendimiento de las estrategias de ajuste individual y grupal</i>	<i>24</i>
<i>1.6.3. Estudio 3. Respecto al rendimiento del modelo jerárquico de rasgo latente en condiciones de heterogeneidad de Matzke et al. (2015).....</i>	<i>26</i>

2. Study 1	29
Abstract	29
2.1. Introduction	29
<i>2.1.1. Doubts and debate about the reality of ICs</i>	31
<i>2.1.2. Explaining conjunction errors</i>	33
<i>2.1.3. Experiment</i>	35
2.2. Method	36
<i>2.2.1. Participants</i>	36
<i>2.2.2. Stimuli and materials</i>	36
<i>2.2.3. Procedure</i>	39
<i>2.2.4. Predictions</i>	39
2.3. Results	41
2.4. Discussion	44
2.5. Some processing models	45
2.6. Conclusions	50
 3. Study 2	 53
Abstract	53
3.1. Introduction	54
<i>3.1.1. Group and individual fitting</i>	56
<i>3.1.2. Multinomial processing tree models</i>	57
<i>3.1.3. Illusory conjunctions</i>	59
3.2. Method	63
<i>3.2.1. Parameters used in the simulation</i>	63
<i>3.2.2. Simulation conditions</i>	65
<i>3.2.3. Data analysis</i>	67
3.3. Results	67
<i>3.3.1. Homogeneity</i>	68
<i>3.3.2. Quantitative heterogeneity</i>	68
<i>3.3.3. Qualitative heterogeneity</i>	69
<i>3.3.4. Effects on the distribution of the estimates</i>	70
3.4. Discussion	73

4. Estudio 3	75
Resumen	75
4.1. Introducción	76
<i>4.1.1. Modelo MPT jerárquico de rasgo latente</i>	76
4.1.1.1. Nivel 1: Modelo MPT	76
4.1.1.2. Nivel 2: Variación individual del vector de parámetros θ_i	77
4.1.1.3. Nivel 3: Matriz de varianzas-covarianzas para los parámetros transformados	78
<i>4.1.2. Modelos MPT para las conjunciones ilusorias</i>	79
4.2. Método	83
<i>4.2.1. Valores paramétricos empleados en la simulación</i>	83
<i>4.2.2. Condiciones en la simulación</i>	84
<i>4.2.3. Análisis de datos</i>	85
4.3. Resultados	86
4.4. Discusión	93
 5. Conclusiones generales	95
5.1. Sugerencias para el ajuste de modelos formales en investigación aplicada	98
5.2. Limitaciones y líneas futuras	99
 5. General conclusions	101
 Referencias	104

Resumen

La forma más frecuente de analizar los resultados en psicología experimental consiste en comparar promedios de variables obtenidas en diferentes grupos o condiciones realizando contrastes de hipótesis de nulidad mediante modelos lineales. Este enfoque implica el supuesto de que los individuos son esencialmente idénticos y solo les diferencian el error de medida asociado a sus mediciones más la presencia de variables extrañas que interesa controlar más que estudiar. Sin embargo, en muchas ocasiones el supuesto de independencia e idéntica distribución (i.i.d.) es muy cuestionable. A pesar de ello, incluso empleando técnicas más sofisticadas como los modelos matemáticos es también habitual asumir implícitamente supuesto de i.i.d. al promediar los resultados por individuos. El problema principal planteado en esta tesis tiene que ver con los efectos de asumir este supuesto en presencia de participantes heterogéneos en la muestra a partir de la que se realiza la evaluación y selección de un modelo formal.

La tesis se compone de tres estudios, uno experimental y dos de simulación. El primer estudio es un trabajo experimental en el que se exploran los errores que se producen en forma de conjunciones ilusorias color/forma en Presentaciones Rápidas de Series Visuales (RSVP). Algunos autores han puesto en duda que los resultados que se obtienen con este paradigma experimental no sean otra cosa que ruido producido por variables confundidas. Los resultados de este experimento refuerzan considerablemente la posición contraria: la de que dichos errores realmente existen. Además, el análisis de los resultados llevó a plantear la posibilidad de que entre la muestra de participantes hubiera algunos que sí forman conjunciones ilusorias y otros que no.

El segundo estudio retoma la cuestión abierta al final del primero, planteando la problemática de la heterogeneidad de los participantes en un experimento en el que se recogen respuestas categóricas, que pueden ser analizadas empleando modelos de árbol. Para ello se realiza un estudio de simulación en el que se generan datos bajo diferentes escenarios de heterogeneidad, distinguiendo entre heterogeneidad cuantitativa bajo un mismo modelo, heterogeneidad cualitativa o de modelos alternativos, y la combinación de los dos anteriores. Naturalmente, como condición de ‘control’ se incluye una de homogeneidad completa. En el estudio se valora el rendimiento de diferentes estrategias analíticas, como el ajuste de datos agregados, promediados o individuales, en las distintas condiciones de heterogeneidad. Entre los principales resultados encontramos

que el ajuste a partir de datos agregados puede enmascarar las condiciones de heterogeneidad cualitativa, llevando a rechazar modelos correctos para gran parte de la muestra y a estimaciones de parámetros sesgadas. Como alternativa, el ajuste individual, aunque ofreciendo estimaciones de parámetros menos precisas, es sensible a situaciones en las que los individuos se ajustan a distintos modelos, mostrando resultados que describen correctamente el comportamiento de toda la muestra de participantes.

El tercer estudio extiende el trabajo anterior a los modelos bayesianos de rasgo latente. En el estudio de simulación que hemos llevado a cabo el rendimiento de los modelos de rasgo latente es similar al obtenido ajustando modelos de árbol tradicionales a cada conjunto individual de datos por separado, permitiendo detectar diferencias discretas entre grupos de participantes a través de distribuciones posteriores con un perfil multimodal. Sin embargo, encontramos que esta ventaja no es tan clara cuando en cada subgrupo existen también diferencias continuas en la probabilidad con la que se emplean los distintos procesos. Los modelos de clase latente presentan también uno de los problemas de la estrategia de ajuste a partir de conjuntos individuales de datos: sobreajuste.

Concluimos con una serie de recomendaciones de cara al trabajo con modelos formales, analizando el impacto que el problema planteado podría estar teniendo en distintas áreas de investigación en psicología cognitiva.

Abstract

The regular way of analyzing results in experimental psychology is comparing mean values of measures obtained in different groups/conditions employing null hypothesis tests by means of linear models. This approach involves assuming that individuals are essentially identical, explaining their observed differences as product of measure error and non-controlled variables. However, in many cases the independent and identical distribution assumption (i.i.d.) is hardly tenable. Nonetheless, even when working with more sophisticated tools - as mathematical models - is also usual to implicitly assume i.i.d. when employing averaged-by-individual data. The main problem addressed in this thesis are the potential negative effects of assuming i.i.d when performing evaluation and selection of formal models with samples of heterogeneous participants.

Three studies are presented in this thesis: one experiment and two simulation studies. The first study is aimed to explore the so-called illusory conjunction errors that can be observed in experiments of Rapid Visual Serial Presentation (RVSP). Some authors questioned the existence of those errors, considering them as mere methodological artifacts. The results obtained in this study, analyzed by fitting a mathematical model, strongly support the opposite: illusory conjunctions truly exist. However, the results suggest the possibility that some participants could show illusory conjunctions while others not.

The second study retakes the open issue of participant heterogeneity in a categorical response framework that can be analyzed by tree models. In this study we simulated data under different conditions of individual heterogeneity, distinguishing between quantitative heterogeneity in presence of a single model, qualitative heterogeneity or presence of alternative models, and both combined. Naturally, a total homogeneity condition is also simulated as the control condition. In this study, performance of different analytical strategies as aggregate, averaged and individual data fit are assessed. The main findings were that fitting models using aggregate data can mask conditions of qualitative heterogeneity, leading to biased parametric estimations and rejection of correct models for a large portion of the sample. Alternatively, the individual fit, though less accurate in parameter estimation, is sensitive to situations in which individuals fit different models, offering results that faithfully describe the behavior of the whole sample of participants.

The third study applies the last work to latent trait bayesian models. In this simulation study the performance of the latent trait model is similar to that yielded by fitting classic tree models with an individual strategy, allowing to detect discrete differences between groups of participants by means of posterior distributions with multimodal profiles. However, this advantage is found to be less clear when there are also continuous differences in the probability represented by parameter values. Latent trait models also show one of the main problems of individual fitting strategy: over-fitting.

The thesis concludes with some recommendations in formal modeling work, analyzing the possible impact of the presented problem in different areas of research in cognitive psychology.

Presentación

El uso del lenguaje natural es una de las formas más comunes de formular y comunicar las teorías en psicología. Entre los objetivos de las teorías se encuentran los de delimitar y simplificar parte de la realidad para comprender o predecir su funcionamiento, pero la falta de precisión de las teorías verbales o informales en muchos casos dificulta esta tarea. En este contexto los modelos matemáticos se presentan como un modo formal de definir teorías con mayor precisión, lo que hace más eficaz tanto su comunicación como la evaluación de su verosimilitud a partir de la evidencia empírica (Botella, 2005; Estes, 1956; Ponsoda, 1986).

Es habitual contrastar la credibilidad de las teorías verbales valorando la magnitud de las diferencias observadas como resultado de la manipulación de un factor propuesto en ellas. En concreto, suele comprobarse si la diferencia es mayor de lo que cabría esperar por mero azar (Botella, 2005; Myung, Pitt & Kim, 2005) en lo que se conoce como *contrastos de significación de hipótesis nulas*. Sin olvidar que tales pruebas estadísticas implican un modelo matemático subyacente (generalmente de carácter lineal), su uso excesivo ha limitado la forma y contenidos que adoptan las teorías, restringiendo su nivel de detalle a contrastes dicotómicos y no de grado (Kirk, 1996). Frente a la evaluación de teorías informales a partir de pruebas generales de significación de hipótesis nulas, el uso de modelos matemáticos ofrece mayor flexibilidad, variedad y especificidad.

El empleo de modelos matemáticos en psicología se remonta a la psicofísica y los trabajos pioneros de Weber sobre los umbrales mínimos para la percepción subjetiva del cambio en magnitudes físicas objetivas. Fechner redescubrió y definió formalmente la relación entre la intensidad objetiva de un estímulo y su percepción subjetiva en la que se refiere habitualmente como la ley de Fechner-Weber (Stevens, 1975). Sin embargo, el uso de modelos formales no empezó a popularizarse en la psicología científica hasta la década de 1950 (Cavagnaro, Myung & Pitt, 2013), haciendo referencia a ellos generalmente como modelos matemáticos. Son ejemplos claros de esta popularización los modelos formales empleados en teoría de la elección (Luce, 1959) o en la teoría de detección de señales (Green & Swets, 1998).

Parte de su popularidad se debe a que la formalización de teorías permite plantear formas más robustas e informativas de poner a prueba postulados teóricos sobre los mecanismos subyacentes al comportamiento (Marewski & Olsson, 2009). Entre otras ventajas, los modelos matemáticos proveen al investigador de herramientas para comparar la idoneidad de distintos mecanismos teóricos propuestos como explicación de un mismo fenómeno comparando el ajuste relativo de sus predicciones a los resultados empíricos. Por otro lado, permiten estudiar la magnitud de los cambios que provocan las manipulaciones experimentales de variables relevantes para los modelos y, por tanto, sobre los procesos cognitivos propuestos. Al implicar mayor precisión y detalle sobre los efectos esperados de distintas manipulaciones, los modelos formales son una herramienta más eficaz para captar la complejidad de un fenómeno psicológico. De forma análoga, la información extraída de los datos experimentales en su análisis al emplear modelos es mucho mayor (Myung, et al., 2005), haciendo más accesible el contraste de teorías explicativas y no sólo descriptivas.

El problema principal planteado en esta tesis tiene que ver con los efectos de la presencia de participantes heterogéneos en la muestra a partir de la que se realiza la evaluación y selección de un modelo formal. Por tanto, comenzaremos la exposición del marco teórico tratando formalmente distintas posibles fuentes de heterogeneidad entre individuos, distinguiendo entre diferencias cuantitativas dentro de un mismo modelo y la eventual presencia de individuos cuyo comportamiento se ajusta a modelos cualitativamente distintos. Continuaremos revisando el efecto de distintas estrategias de tratamiento de los datos en presencia de estas fuentes de diferencias así como distintos modelos formales propuestos para dar cuenta de ellas. A continuación describiremos brevemente las conjunciones ilusorias, el fenómeno psicológico cuya evidencia empírica emplearemos para ilustrar nuestras propuestas en el terreno de los modelos matemáticos. Después plantearemos nuestra conceptualización de dicho fenómeno en términos de modelos multinomiales de procesamiento en árbol. Terminaremos la introducción exponiendo el plan general de trabajo y presentando tres capítulos con un estudio experimental y dos de simulación. Por último, propondremos unas conclusiones generales basadas en los resultados obtenidos en los estudios.

Es fácil advertir en la estructura de este documento que el plan original era presentar una tesis como compendio de publicaciones. No ha sido posible terminarla a tiempo para presentarla formalmente en ese formato dentro del plazo exigido, pues en el

momento de cerrarla aun nos faltaba la aceptación de dos de los tres artículos. Sin embargo, hemos preferido conservar la estructura que hubiera tenido de haber llegado a tiempo, aunque formalmente sea una tesis en el formato convencional y no una basada en publicaciones.

1. Introducción

1.1. Posibilidades no contempladas en el ajuste a partir de datos agregados: el supuesto de independencia e idéntica distribución (i.i.d.)

Al analizar datos experimentales, una de las principales ventajas de modelar formalmente procesos cognitivos respecto a una aproximación basada en técnicas clásicas de uso general (e.g. ANOVA) es una mayor flexibilidad y precisión en el desarrollo y evaluación de teorías. Los modelos formales de procesos cognitivos pueden centrarse en la descripción y/o explicación de: (1) Procesos cognitivos básicos que todos los individuos comparten o (2) diferencias entre individuos, o grupos de ellos en algún aspecto de su funcionamiento cognitivo.

Quizá porque el primer enfoque es, con mucho, más frecuente en contextos de investigación experimental, también lo es asumir el supuesto de *independencia e idéntica distribución* (i.i.d.) entre participantes en las variables bajo estudio. Aunque en ocasiones no se haga de forma explícita, es frecuente asumir i.i.d. entre individuos como consecuencia de abordar la estimación y ajuste de modelos formales a partir de datos agregados o promediados. Cuando el supuesto de i.i.d. se cumple, todos los individuos se ajustan a un único modelo con valores fijos para sus parámetros θ :

$$g(\mathbf{y}) = f_{\text{datos}}(\mathbf{y} | \theta), \quad (1.1)$$

donde \mathbf{y} es el vector de datos y f_{datos} es la función de distribución con la que predecir el vector de resultados \mathbf{y} a partir del modelo formal propuesto. El vector \mathbf{y} puede ser un vector de variables discretas, siendo el caso para el que están pensados, entre otros, los modelos multinomiales de procesamiento en árbol (MPT).

A pesar de que el estudio de procesos cognitivos generales es más común, puede haber tareas experimentales en las que el desempeño de los participantes esté basado en fenómenos cognitivos heterogéneos. Esta heterogeneidad puede deberse a:

(a) *Diferencias continuas entre participantes que siguen el mismo modelo.* Es una situación en la que todos los participantes siguen un mismo modelo, pero existe un vector de parámetros θ_i para cada individuo i . Los valores de los parámetros del modelo varían entre individuos de acuerdo a una función de densidad de probabilidad $f_{\text{parámetros}}$:

$$g(\mathbf{y}) = \int f_{\text{datos}}(\mathbf{y} | \boldsymbol{\theta}_i) f_{\text{parámetros}}(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i. \quad (1.2)$$

(b) *Diferencias discretas entre participantes que siguen el mismo modelo.* Se trata de un caso particular de (a) en el que los valores de los parámetros del modelo están definidos en $p > 1$ puntos de acuerdo a P subpoblaciones distintas que se ajustan al mismo modelo f_{datos}

$$g(\mathbf{y}) = \sum_{p=1}^P \alpha_p \cdot f_{\text{datos}}(\mathbf{y} | \boldsymbol{\theta}_p), \quad (1.3)$$

donde α_p es la proporción p de participantes que se ajustan al modelo f_{datos} con un vector de parámetros $\boldsymbol{\theta}_p$.

(c) *Diferencias discretas entre participantes debidas a modelos distintos.* Un último posible escenario es aquel en el que los participantes pertenecen a m subpoblaciones cuyos resultados se ajustan a $z > 1$ modelos $f_{z,\text{datos}}$ que describen distintos procesos cognitivos:

$$g(\mathbf{y}) = \sum_{z=1}^m \alpha_z \cdot f_{z,\text{datos}}(\mathbf{y} | \boldsymbol{\theta}_z), \quad (1.4)$$

donde α_z es la proporción de participantes que se ajusta al z -ésimo modelo en la población.

Es importante mencionar que las situaciones (a), (b) y (c) no son mutuamente excluyentes. De encontrar mayor variabilidad en los datos obtenidos en un experimento de la esperable por mero muestreo, la sobredispersión puede deberse a una de las fuentes mencionadas o a la mezcla de cualquiera de ellas. Por ejemplo, pueden existir m subpoblaciones que se ajustan a distintos modelos cuyos parámetros varían entre individuos de acuerdo a una o varias f.d.p. $f_{z,\text{parámetros}}$

$$g(\mathbf{y}) = \sum_{z=1}^m \int \alpha_z \cdot f_{z,\text{datos}}(\mathbf{y} | \boldsymbol{\theta}_i) f_{z,\text{parámetros}}(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i. \quad (1.5)$$

Estas situaciones son una fuente de heterogeneidad plausible en auténticos contextos experimentales (Rouder & Lu, 2005) y, en concreto, la situación descrita en (c) solo ha sido planteada recientemente (ver Speelman & Muller-Townsend, 2015; Houpt, Yang & Townsend, 2016). Dado que podrían dar lugar a sesgos y pérdida de capacidad para generalizar aportaciones a la teoría cognitiva, es importante conocer cómo afectan a los procesos de estimación de parámetros y ajuste de modelos formales.

En este trabajo nos proponemos estudiar el efecto de las situaciones de heterogeneidad propuestas sobre la evaluación de modelos formales según: (1) las estrategias de tratamiento de los datos previas a la estimación y ajuste de modelos; y (2) el empleo de modelos específicamente diseñados para dar cuenta de la sobredispersión.

1.2. Estrategias de tratamiento de datos previas a la estimación y ajuste.

Ventajas y limitaciones

La estrategia de tratamiento del conjunto de datos a partir de los que realizar la estimación y ajuste de modelos formales puede tener una fuerte influencia sobre los resultados obtenidos en las condiciones de heterogeneidad entre participantes mencionadas en el apartado anterior. Las estrategias generalmente propuestas y más estudiadas son (1) ajuste a partir de datos agrupados y (2) ajuste a partir de datos individuales.

1.2.1. Datos agrupados y estimación de un conjunto de parámetros general

Realizar la estimación de parámetros y el ajuste de un modelo a partir de datos agrupados por participantes supone asumir que no existen diferencias individuales. De cumplirse el supuesto de i.i.d. se trata de la mejor estrategia a adoptar, ya que permite describir y predecir correctamente el comportamiento de los individuos con el menor número de parámetros posible: un único vector θ general. Por otra parte, esta simplicidad evita problemas de sobreajuste (Lee & Webb, 2005), que pueden aparecer cuando los modelos más complejos muestran mejor ajuste adaptándose a variaciones no sistemáticas (y que, por tanto, no se busca explicar) entre participantes. Cuando esto ocurre los resultados obtenidos con distintas muestras son poco consistentes, lo que redundará en menores garantías de generalización de resultados a la población de referencia.

Sin embargo, en caso de existir diferencias sistemáticas entre participantes estas podrían quedar enmascaradas al emplear datos agrupados por individuos, especialmente si el modelo empleado es de carácter no lineal (Estes, 1956; Rouder & Lu, 2005). Por un lado, la variabilidad de los parámetros estimados podría quedar infraestimada; por otro, se corre el riesgo de obtener estimaciones que a pesar de ofrecer el mejor ajuste

posible para el modelo especificado son poco representativas del desempeño de los participantes.

1.2.2. Datos individuales y estimación de un conjunto de parámetros por participante

Una de las estrategias empleadas para superar las limitaciones del ajuste a partir de datos agrupados es emplear los datos individuales y estimar tantos vectores de parámetros θ_i distintos como participantes, donde $i = \{1, \dots, i, \dots, N\}$ y N es el tamaño muestral. Siguiendo esta estrategia no existe el riesgo de obtener estimaciones sesgadas en presencia de diferencias sistemáticas entre participantes: ni se subestimaría la variabilidad entre parámetros ni se enmascararía la presencia de grupos de participantes con un desempeño distinto.

A pesar de sus ventajas, emplear esta estrategia supone un riesgo de sobreajuste del modelo a las particularidades del conjunto de datos empleado para evaluarlo. Así, las estimaciones obtenidas pueden ser menos precisas dada su mayor sensibilidad a la presencia de variación aleatoria en los datos, que sistemáticamente se atribuiría a diferencias entre participantes (Navarro, Griffiths, Steyvers & Lee, 2006). Por otra parte, emplear esta estrategia supone trabajar con un número mucho mayor de parámetros, lo que supone relajar el compromiso con el principio de parsimonia.

1.2.3. Estrategia mixta: grupos de participantes distintos

Una estrategia más flexible para abordar los posibles casos de violación de i.i.d. es la propuesta en Lee & Webb (2005) o Navarro et al. (2006). Consiste en proponer una familia de modelos en la que tanto el caso de un único vector de parámetros θ para todos los individuos como tantos vectores θ_i como participantes son casos particulares y extremos. Siguiendo esta estrategia, se especificarían P subgrupos, (donde $P = \{1, \dots, p, \dots, N\}$) cada uno con sus propio vector de valores paramétricos θ_p , que se estimarían a partir de datos agrupados únicamente entre los individuos homogéneos entre sí. Idealmente, esta estrategia se centraría en seleccionar el modelo que mostrara mejor ajuste especificando el menor número P de subgrupos posible. El modelo escogido siguiendo estos criterios sería el más informativo, ya que con él se explicarían tanto las regularidades como las diferencias entre participantes. Con esta estrategia, además, se

pretende evitar sesgos en las estimaciones sin aumentar innecesariamente la complejidad del modelo y, por tanto, descartar el riesgo de sobreajuste a la variabilidad no sistemática presente en los datos.

1.3. Modelos propuestos para formalizar diferencias individuales

Expondremos la estructura básica de los modelos propuestos para tener en cuenta situaciones de heterogeneidad entre participantes distinguiendo entre aquellos planteados para abordar (1) diferencias continuas en los valores de los parámetros y (2) diferencias discretas.

1.3.1. Modelos para diferencias continuas: efectos aleatorios y rasgo latente

Un modo de formalizar la situación de diferencias continuas en los parámetros de un modelo que siguen todos los participantes es especificar distintos niveles jerárquicos. Un primer nivel describiría el comportamiento del grupo, y a partir de él se formalizarían las diferencias entre individuos en un segundo nivel jerárquico. En el nivel grupal se describiría una estructura común a todos los participantes, con un vector de parámetros θ . Por su parte, el nivel individual permitiría tener en cuenta la variación en el vector de parámetros θ_i para los N participantes de acuerdo a una función de densidad de probabilidad. Este sería el caso de los modelos de efectos aleatorios y jerárquicos de rasgo latente.

En estos modelos la función de densidad de probabilidad empleada en el nivel individual se escoge frecuentemente de entre la familia Gaussiana, con lo que se asume que la variación se concentra en torno a un único valor central (i.e. unimodalidad), que se corresponde con el parámetro del grupo. Cuando los participantes se ajustan al mismo modelo con parámetros agrupados en torno a más de un valor (ecuación 1.3) o cuando se ajustan a distintos modelos (ecuación 1.4), el supuesto de unimodalidad en el nivel individual podría enmascarar la presencia de distintos grupos de participantes.

Dentro de la familia de los modelos multinomiales de procesamiento en árbol (MPT), que trataremos en detalle más adelante, los modelos beta-MPT (Smith & Batchelder, 2010) y los modelos jerárquicos de rasgo latente (Klauer, 2010; Matzke, Dolan, Batchelder & Wagenmakers, 2015) pertenecen a esta categoría. En ellos se

formaliza la variación individual en los valores de los parámetros empleando una función beta y una función normal multivariada, respectivamente.

1.3.2. Modelos para diferencias discretas: clase latente

Existen propuestas teóricas menos restrictivas respecto a la forma en la que se entiende que se podrían distribuir las diferencias entre participantes, especialmente respecto a una naturaleza no necesariamente continua y/o unimodal. Éstas se basan en plantear un número finito (Lee & Webb, 2005) o potencialmente infinito (Navarro et al., 2006) de vectores de parámetros θ_p a estimar. Cada vector θ_p se correspondería con un subgrupo de participantes clasificados dentro de un mismo grupo. Para cada subgrupo con resultados similares se estimaría, además, su peso específico dentro de la población.

Esta aproximación permitiría describir la multimodalidad que puede aparecer en las distribuciones de los parámetros cuando existen diferencias discretas entre los participantes (situación b) definiendo no sólo diferencias entre grupos de individuos sino *similitudes* entre ellos en los procesos cognitivos empleados en la resolución de tareas experimentales.

A pesar de que esta fuente de variabilidad individual ha recibido menos atención, entre los modelos de árbol el modelo de clase latente de Klauer (2006) pertenecería a este grupo.

En esta tesis empleamos el debate sobre las conjunciones ilusorias (Botella, Barriopedro & Suero, 2001) y los modelos de árbol propuestos para formalizar las distintas posturas planteadas en él como ejemplo sustantivo. Tomando este caso como ejemplo estudiaremos el efecto de distintas fuentes de heterogeneidad entre individuos en los resultados de la evaluación de modelos formales. Por ello, en los siguientes apartados trataremos la controversia sobre las conjunciones ilusorias, seguida de una descripción formal de los modelos de árbol.

1.4. Atención y conjunciones ilusorias

En experimentos enfocados al estudio de procesos atencionales se ha observado que los participantes en ocasiones cometen errores consistentes en reportar

combinaciones de características realmente presentadas, aunque no como parte de un único estímulo (e.g. reportan una R azul cuando realmente se han presentado una T azul y una R roja). Este fenómeno ha sido importante para la fundamentación de la teoría de integración de rasgos o *Feature Integration Theory* (FIT; Treisman & Gelade, 1980; Treisman & Schmidt, 1982), haciendo referencia a él con el término *conjunciones ilusorias* (CI).

La primera versión de la FIT proponía dos etapas secuenciales en el procesamiento de estímulos visuales. En la primera etapa se extraen las características básicas del estímulo como la forma o el color. Estas características son procesadas en paralelo y a nivel preatencional, generando lo que llamaron “rasgos flotantes”. Estos rasgos flotantes son representaciones descoordinadas de características de un estímulo (e.g., forma circular y color azul, por separado). En la segunda etapa se forma una representación visual integrada o percepto combinando las características simples extraídas en la etapa anterior (en el ejemplo, un círculo azul). En esta segunda etapa los perceptos se procesan de un modo serial y controlado, lo que requiere la intervención de la atención focal. Dado que la atención focal es el mecanismo que “pega” los rasgos flotantes, si el proceso se ve entorpecido de algún modo es posible que los rasgos flotantes se combinen de forma incorrecta.

A pesar de que la FIT ha ido sufriendo revisiones (Quinlan, 2003), sigue defendiendo que: (a) los rasgos flotantes existen y (b) son propensos a combinarse de forma incorrecta cuando se desvía la atención focal, dando lugar a CI reales. La existencia de las CI es uno de los soportes principales de la teoría, esencialmente porque apoya una arquitectura en la que existe una etapa previa a la formación de los perceptos en la que existen representaciones de características básicas.

Como fenómeno congruente con la arquitectura del modelo, la FIT fue el primer modelo que recurría a la existencia de CI como fuente de evidencia. Más tarde otros modelos, como el de Botella et al. (2001) también han supuesto la existencia de representaciones de características básicas aisladas. Si la evidencia respecto a la existencia de las CI se pusiera en entredicho también lo haría parte del apoyo empírico que sustenta la FIT y modelos similares.

El término CI puede llevar a error dado que en vez de ser descriptivo implica asumir el modo en que desde la FIT se entiende que el sistema cognitivo produce los errores mencionados. A un nivel meramente descriptivo existen los *errores de*

conjunción, que son respuestas en las que se combinan las características de dos estímulos realmente presentados. En las mismas tareas experimentales estos errores aparecen junto con otros tipos (e.g. reportar un estímulo que incluye una característica no presentada).

Los errores de conjunción no tienen por qué ser entendidos como CI. Por ejemplo, si un participante elige una característica de forma aleatoria para responder, esta podría coincidir con una de las presentadas como estímulo. Del mismo modo, pueden darse errores de conjunción que sean el resultado de confusiones en la categorización. Por ejemplo, cuando el observador debe reportar la única letra roja presentada y confunde la forma de esa letra roja. Si la letra roja presentada es una T pero se categoriza incorrectamente como una L habiendo presentado también una L, el observador reportaría una letra presentada. Así, tanto las respuestas dadas al azar como los fallos en la categorización son procesos distintos a las CI que también pueden producir errores de conjunción.

Precisamente Donk (1999, 2001) destaca estos otros procesos para poner en cuestión la posición que entiende los errores de conjunción como una prueba de la existencia de las CI. Para ella, al poder explicarse a partir de otros procesos, la existencia de las CI no ha sido probada, al menos entendidas como producto de una combinación incorrecta de características correctamente identificadas. Así, para Donk los errores de conjunción son artefactos experimentales que pueden explicarse sin necesidad de recurrir a las CI. Por su parte, Ivry & Prinzmetal (1991) encontraron que los errores de conjunción son más probables para características más similares entre estímulos, lo que apoya la idea de que al menos parte de ellos se producen por categorizaciones incorrectas.

Sin embargo, algunos autores defienden que incluso teniendo en cuenta y descontando estas fuentes alternativas de errores de conjunción, algunos de estos errores aún pueden considerarse verdaderas CI (Ashby, Prinzmetal, Ivry & Maddox, 1996; Prinzmetal, Diedrichsen & Ivry, 2001).

Pensamos que esta controversia puede haber ocurrido por dos razones:

La primera es que la principal fuente de evidencia son las CI que ocurren en el dominio espacial, es decir, entre características pertenecientes a objetos presentados de forma simultánea en diferentes localizaciones espaciales. Este tipo de CI es infrecuente.

En cambio, las CI que se dan en un dominio temporal son un fenómeno mucho más potente y frecuente. Lawrence (1971) fue el primero en describirlas, y Botella y colaboradores han publicado numerosos experimentos mostrando migraciones de características presentadas en momentos distintos (e.g., Botella & Eriksen, 1992; Botella, Garcia & Barriopedro, 1992; Botella, Narvaez, Suero & Juola, 2007) además de proponer un modelo que busca explicar el modo en que se forman (Botella et al., 2001).

La segunda posible razón para la controversia es que al analizar la evidencia no se ha tenido en cuenta que en la mayoría de los ensayos en los que los participantes no cometen errores esto ocurre porque la atención se focaliza adecuadamente en dicho estímulo. Que esto ocurra depende de factores como el tiempo en que cada estímulo permanece presente (SOA o *stimulus onset asynchrony*), el ángulo visual, la discriminabilidad del estímulo o la magnitud del intervalo entre estímulos en la presentación de series visuales. De existir, los mecanismos propuestos para dar cuenta de los errores solo se pondrían en marcha al fallar la atención focal. Así como las propiedades de cada condición experimental determinan la probabilidad de un funcionamiento adecuado de la atención focal, también influyen sobre la probabilidad de cualquier otro mecanismo propuesto como causa de los errores de conjunción (respuestas basadas en rasgos flotantes, errores de categorización, respuestas al azar, etc). Por ejemplo, Botella & Eriksen (1991) mostraron que la frecuencia de errores de conjunción aumenta a medida que se reduce el SOA.

1.4.1. El paradigma de presentación rápida de series visuales

Los estudios experimentales sobre CI habitualmente emplean el paradigma de presentación rápida de series visuales (PRSV). En los experimentos que emplean este paradigma se presentan en una pantalla series de 15-20 letras coloreadas¹. Al comienzo de cada ensayo se presenta una letra para indicar cuál es la letra *target* en ese ensayo. El participante elige cuando comenzar la serie, momento a partir del cual cada letra permanece en la pantalla durante un breve espacio temporal. Tras terminar la secuencia,

¹Aunque la explicación se desarrolle haciendo referencia a la presentación de letras y colores, en un experimento de PRSV pueden emplearse otras dimensiones estimulables

el participante debe reportar el color en el que apareció la letra *target*. La alta velocidad de presentación impide percibir todos los estímulos presentados, de modo que el porcentaje de aciertos suele darse en torno al 60%. Por su parte, los errores suelen consistir en reportar un color perteneciente a un estímulo presentado y cercano en el tiempo al *target* (ver Botella & Eriksen, 1992).

En este trabajo empleamos un modo de presentar los estímulos que permite distinguir las CI que ocurren en el dominio espacial de aquellas que se dan en el temporal. Las respuestas de los participantes pueden clasificarse en cinco categorías basándose en su relación con el estímulo *target*: (1) aciertos, cuando coinciden con él, (2) intrusiones de estímulos presentados en la misma localización espacial pero en distinto momento, (3) intrusiones de estímulos presentados en distinta localización pero en el mismo momento, (4) estímulos que no han aparecido presentados, (5) omisiones o ausencia de respuesta.

Consideramos dos formas generales de concebir cómo y por qué aparecen las intrusiones entre los resultados de tareas experimentales de PRSV: aquellas que apoyan la existencia de representaciones de características aisladas en el sistema cognitivo y las que no. Entendemos que para contrastar estas posiciones teóricas de forma sustantiva y completa, además del contraste de predicciones ordinales es deseable emplear modelos formales. En concreto, emplearemos modelos multinomiales de procesamiento en árbol, que describiremos a continuación.

1.5. Modelos multinomiales de procesamiento en árbol (MPT)

Los modelos multinomiales de procesamiento en árbol (*multinomial processing tree* o MPT) han sido desarrollados para modelar procesos cognitivos en contextos de respuestas categóricas. Permiten formalizar y poner a prueba la verosimilitud de suposiciones sobre los mecanismos cognitivos subyacentes a resultados experimentales (Riefer & Batchelder, 1988). Formalmente, los modelos MPT o modelos de árbol reparametrizan las probabilidades π_j asociadas a un conjunto de $j = \{1, \dots, j, \dots, J\}$ categorías discretas C_j contempladas en un paradigma experimental específico. Se asume que el número de ensayos n_j en los que el resultado coincide con cada categoría C_j , sigue una distribución multinomial

$$f(\mathbf{x}; n, \boldsymbol{\pi}) = n! \prod_{j=1}^J \pi_j^{n_j} / n_j!, \quad (1.6)$$

donde n hace referencia al número de ensayos total, por lo que $n = \sum n_j$; \mathbf{x} es un vector de frecuencias formado por los valores de n_j y $\boldsymbol{\pi}$ es el vector con los distintos valores de π_j .

La estructura del modelo de árbol especifica S estados cognitivos subyacentes a la observación de cada categoría observable C_j . Basándose en estos estados subyacentes, las ecuaciones del modelo de árbol especifican “rutas de procesamiento” o secuencias de estados cognitivos. El vector de parámetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_s, \dots, \theta_S)$ del modelo hace referencia a la probabilidad de que cada uno de estos estados tenga lugar. Así, en un modelo de árbol los valores de las probabilidades π_j son función del vector $\boldsymbol{\theta}$.

El modelo especifica un número B_j de rutas r_{bj} en las que se suceden secuencialmente un conjunto de estados cognitivos. En cada ruta pueden verse involucrados unos estados cognitivos, cada uno con probabilidad θ_s , pero también la ausencia de otros, en cuyo caso la probabilidad se obtendría a partir del complementario $1 - \theta_s$. Cada ruta da lugar a una categoría C_j , aunque cada categoría puede especificarse como resultado de una o varias rutas, según la estructura del modelo. Formalmente, la probabilidad para cada ruta propuesta por el modelo sería

$$P(r_{bj}; \boldsymbol{\theta}) = c_{bj} \prod_{s=1}^S \theta_s^{a_{bjs}} \cdot (1 - \theta_s)^{u_{bjs}}; 0 \leq \theta_s \leq 1, \quad (1.7)$$

donde c_{bj} es el producto de constantes positivas asociadas a los estados y a_{bjs} y u_{bjs} son enteros no negativos que recogen el número de ocasiones en que la ruta involucra un proceso cognitivo s (con probabilidad θ_s) o su complementario (con probabilidad $1 - \theta_s$) (Hu & Batchelder, 1994). Únicamente en casos especiales ocurre que $c_{bj} \neq 1$, como los modelos multinomiales conjuntos, propuestos para varios sistemas de categorías de respuesta (ver Riefer & Batchelder, 1988), que no trataremos aquí.

Una vez especificadas las rutas del modelo, puede obtenerse la predicción de la probabilidad π_j para cada categoría como la suma de probabilidades de las ramas que desembocan en dicha categoría

$$\pi_j(\boldsymbol{\theta}) = \sum_{b_j=1}^{B_j} P(r_{bj}; \boldsymbol{\theta}). \quad (1.8)$$

Antes de tratar los modelos de árbol con los que formalizar las posiciones teóricas respecto a la existencia de las CI, definiremos la forma de categorizar las respuestas dentro del paradigma de PRSV cuyas probabilidades pretenden predecir. Después, trataremos las propuestas teóricas que representa cada modelo de árbol y los procesos a los que hacen referencia los parámetros que incluyen.

1.5.1. Categorías de respuesta en el paradigma PRSV

En un experimento como el que describiremos en el capítulo 2 (Botella, Suero & Durán, 2017) cada participante completa n ensayos en los que emite una respuesta que, en función de su relación con el estímulo *target*, pertenece a una categoría j de entre las J posibles. Si X_{ij} es la frecuencia con que las respuestas del participante i pertenecen a la categoría j , entonces $\sum_{j=1}^J X_{ij} = n$, donde n se asume constante para todos los participantes.

Para modelar los resultados obtenidos con este paradigma, X_{ij} se considera una variable aleatoria discreta perteneciente a un vector de variables aleatorias \mathbf{X}_i . Este vector sigue una distribución multinomial con parámetros $(n, \boldsymbol{\pi}_i)$, donde $\boldsymbol{\pi}_i$ es el vector de probabilidades de las categorías de respuesta para un participante dado. Este vector incluye seis parámetros:

- $\pi_{Acierto}$: Recoge la probabilidad de que el estímulo reportado coincida con el *target*.
- π_{IT} : Es la probabilidad de que ocurra una intrusión de estímulos presentados en la misma localización que el *target*, pero en un momento temporal distinto en un ensayo dado.
- π_{IE} : Hace referencia a la probabilidad de reportar un estímulo presentado en una localización horizontal respecto al punto de fijación distinta a aquella en la que apareció el *target*. Dentro de esta categoría solo se tienen en cuenta estímulos presentados junto al *target* (en el mismo momento temporal).
- π_{IV} : Se refiere a la probabilidad de que se dé una intrusión desde un estímulo presentado en una localización vertical al punto de fijación. Además de su localización, los estímulos presentados en la vertical tienen la particularidad de

que el participante sabe que en ellas no puede aparecer el estímulo *target*, por lo que no son localizaciones explícitamente atendidas.

- π_{Imp} : Es la probabilidad de responder con un estímulo importado, no presentado en ese ensayo.
- π_{Om} : Recoge la probabilidad de no dar respuesta alguna u omisión.

1.5.2. Modelos de árbol para formalizar las posiciones sobre la existencia de las CI

En los estudios presentados en esta tesis empleamos varios modelos de árbol con los que dar cuenta de las categorías de respuesta contempladas en el paradigma de PRSV. El modo en que dichos modelos formalizan los procesos que propone cada uno de ellos es el siguiente (ver Figura 1.1):

- *Modelo 1 (modelo nulo). Errores de conjunción fruto de procesos no sustantivos: las CI no existen.* Formaliza la posición que explica los errores de conjunción exclusivamente como fruto de respuestas al azar o categorizaciones incorrectas. El modelo define la probabilidad de que se den $S = 2$ procesos, representados por los parámetros F y g . El parámetro F cuantifica la probabilidad de focalización adecuada de la atención en el estímulo *target*, lo que produciría siempre aciertos. Por su parte, el parámetro g daría cuenta de las respuestas no basadas en rasgos flotantes que seguirían a un fallo en la focalización de la atención en el estímulo *target*, lo que puede dar como resultado respuestas pertenecientes a cualquier categoría, exceptuando omisiones. Este último parámetro daría cuenta tanto de los errores de categorización como de las respuestas al azar. Las respuestas producidas por ambos procesos se distribuirían igual entre las categorías de respuesta, haciéndolos indistinguibles. La razón es que el rasgo reportado en el caso de confundirlo con el realmente presentado puede aparecer con la misma probabilidad en cualquier posición, siendo así en todos los ensayos de la tarea.
- *Modelo 2. Rasgos flotantes procedentes de cualquier fuente.* Postula que las CI existen y pueden formarse con la misma probabilidad a partir de rasgos de estímulos presentados en cualquier localización. Obviamente, esto excluye las características no presentadas. Mientras las verdaderas CI deben formarse con rasgos de estímulos realmente presentados, las respuestas no basadas en rasgos

flotantes pueden incorporar rasgos tanto de estímulos presentados como no presentados. Con respecto al anterior, éste modelo incluye el parámetro R , que cuantifica la probabilidad de dar una respuesta basada en rasgos flotantes tras un fallo en la atención focal. En este caso, las respuestas que pueden deberse a R son los aciertos, IT, IE e IV.

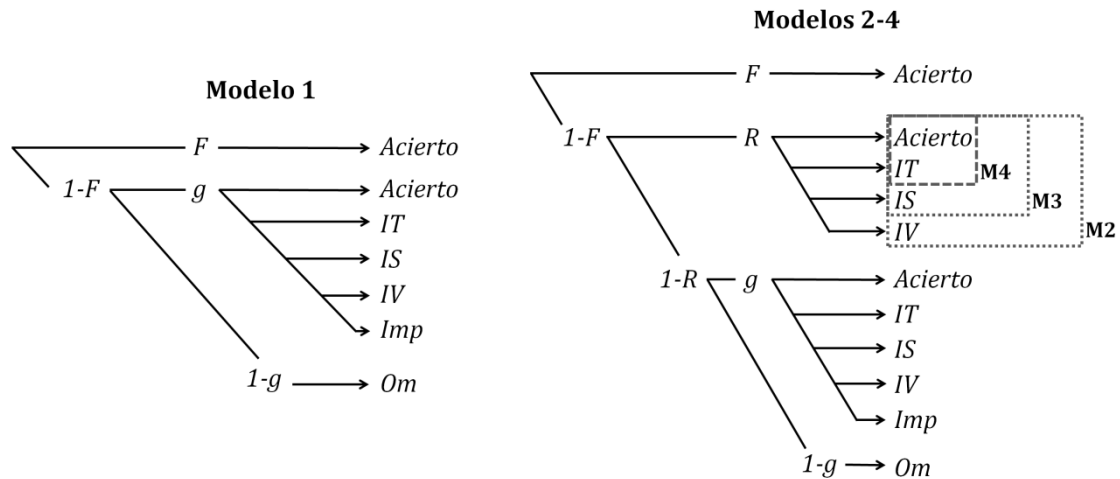


Figura 1.1. Representación gráfica de las rutas propuestas por los modelos de árbol para las categorías de respuesta del paradigma PRSV.

- *Modelo 3. Rasgos flotantes procedentes de localizaciones atendidas.* Defiende que las CI existen, pero solo como migraciones entre los rasgos de estímulos presentados en las localizaciones a las que el observador atiende. Se entienden por localizaciones atendidas aquellas en las que el observador espera que aparezca el estímulo *target*. La ruta que simboliza el parámetro R en este caso sólo puede dar lugar a aciertos, IT e IE entre todas las categorías posibles.
- *Modelo 4. Rasgos flotantes en una dimensión temporal.* Formaliza la posición teórica que apoya la existencia de CI únicamente en el dominio temporal. Desde este modelo se entiende que las características aisladas disponibles para dar una respuesta proceden de estímulos presentados en la misma localización espacial que el estímulo *target*, aunque en momentos distintos. La ruta representada por el parámetro R en este caso únicamente puede dar lugar a respuestas dentro de las categorías de acierto o IT.

1.6. Hipótesis y plan general de estudios

En los siguientes apartados se desglosan los objetivos que se pretenden abordar en este trabajo, mostrando una aproximación general de la estrategia y plan de obtención de evidencia que siguen los estudios incluidos.

1.6.1. Estudio 1. Respecto al debate sobre las CI

El objetivo de este estudio es demostrar que las CI existen incluso teniendo en cuenta los errores de conjunción que pueden deberse a otras causas. En muchos de los experimentos con los que se pretendía poner a prueba la existencia de las CI, éstas eran difíciles de detectar porque el paradigma solo permitía su aparición en una dimensión espacial. Para superar esta limitación, en el Estudio 1 empleamos un paradigma experimental que hace posibles las CI tanto en una dimensión temporal como espacial. De acuerdo a la literatura previa, esperamos encontrar una fuerte evidencia a favor de la existencia de las CI en un dominio temporal, así como una evidencia más débil respecto a las CI en un dominio espacial.

El paradigma experimental empleado consiste en la presentación de series de letras coloreadas en varias posiciones espaciales. La tarea de los participantes consiste en reportar la letra que ha aparecido en un color previamente especificado. Empleando este paradigma se pueden calcular las frecuencias esperadas de respuestas coincidentes con las letras presentadas en todas las posiciones espaciales bajo el supuesto de que todos los errores de conjunción se deben a errores de clasificación o a respuestas al azar. Así, cualquier alejamiento en los resultados respecto a esta predicción requeriría una explicación alternativa, como la existencia de CI.

Desde la hipótesis de aleatoriedad en los errores de conjunción también pueden hacerse predicciones sobre la frecuencia relativa con la que se reportan letras presentadas en la misma localización espacial que el *target*, presentadas en posiciones espaciales distintas o letras no presentadas. Encontrar que la frecuencia relativa de errores que provienen de estímulos realmente presentados es mayor de la esperada bajo esta hipótesis supondría una evidencia a favor de la representación de características aisladas en el sistema cognitivo. De encontrar resultados favorables a la existencia de las CI, pondremos a prueba distintas hipótesis relativas a sus posibles fuentes a partir

del ajuste de modelos que podrían situarse entre la hipótesis de aleatoriedad (las CI no existen) y los enunciados básicos de la FIT (todas las características presentadas tienen el mismo estatus como potenciales respuestas).

En concreto, esperamos encontrar claras desviaciones respecto a lo esperado bajo la hipótesis de aleatoriedad. Además, estas desviaciones deberían ser mayores entre estímulos presentados en la misma localización que el *target* (atribuibles a CI en el dominio temporal) que entre aquellos presentados en distinta localización en el mismo momento (atribuibles a CI en el dominio espacial).

En este primer estudio empírico quedarán patentes las ventajas del ajuste de modelos matemáticos (en este caso, modelos de árbol) sobre el contraste de hipótesis nulas tradicional. La respuesta a la pregunta sobre la existencia de conjunciones ilusorias será más clara y contundente, incluso empleando un ajuste a las frecuencias medias. Este estudio nos servirá para formular algunos modelos en un contexto sustantivo de la psicología cognitiva. También nos servirá para ilustrar la cuestión central de esta tesis. Es posible que el ajuste a un modelo general esté ocultando diferencias individuales, ya sean cuantitativas o de modelos. Es más que probable que los individuos que se ajustan a un mismo modelo no lo hagan con los mismos parámetros. También es plausible que en esta muestra haya individuos que actúan atencionalmente según se describe en dos o más de los modelos formulados. Todas estas cuestiones nos llevan a realizar los dos estudios mencionados a continuación.

1.6.2. Estudio 2. Respecto al rendimiento de las estrategias de ajuste individual y grupal

El Estudio 2 se centra en la posibilidad de que en una misma muestra pudieran encontrarse tanto participantes que siguieran el Modelo 1 como participantes que siguieran el Modelo 4². Desde la perspectiva del modelo del *zoom* (Castiello & Umiltà, 1990; Eriksen & St. James, 1986; Eriksen & Yeh, 1985; LaBerge, Carlson, Williams & Bunney, 1997) es plausible que algunos individuos empleen un foco atencional estrecho

²A partir de este estudio solo se emplean dos de los modelos MPT presentados (los modelos 1 y 4). Por ello, en los estudios 2 y 3 se hará referencia a ellos como modelos 1 y 2

que se orientaría a la localización espacial donde aparece la característica definitoria del estímulo *target*. Esto aumentaría la probabilidad de una focalización efectiva, pero reduciría la disponibilidad de rasgos flotantes en el sistema a los que recurrir cuando la focalización falla. Como contrapartida, otros individuos podrían emplear un foco atencional amplio. Esta segunda estrategia reduciría la probabilidad de una focalización adecuada de la atención en el estímulo *target*, pero generaría abundantes rasgos flotantes en los que basar la respuesta en las ocasiones en las que el primer proceso fallara (Botella et al., 2001).

Atendiendo a esta posibilidad, en una misma muestra de participantes se podrían dar tres escenarios: que solo hubiera participantes que emplearan un foco atencional estrecho, ajustándose sus resultados a la predicción del Modelo 1; que los participantes emplearan únicamente un foco atencional amplio, ajustándose al Modelo 2 o que hubiera tanto participantes que emplearan un foco atencional estrecho como un foco atencional amplio. De acuerdo a las fuentes de heterogeneidad descritas previamente, en las dos primeras situaciones se cumpliría el supuesto de i.i.d., mientras que la última sería un caso de diferencias discretas entre participantes debidas a modelos distintos.

Por otra parte, también puede darse el caso de que entre los participantes que emplean una misma estrategia de foco atencional la frecuencia con la que, por ejemplo, focalizan la atención de forma adecuada varíe. De ocurrir esto, existiría variabilidad en los valores paramétricos de los modelos desde los que se describe su comportamiento. Si esta variabilidad es continua, los parámetros del modelo se podrían concebir como variables aleatorias que siguen una distribución desconocida, lo que se correspondería con la situación de heterogeneidad entre individuos previamente descrita como diferencias continuas entre participantes que siguen el mismo modelo.

En el Estudio 2 se simulan los resultados que podrían obtenerse en experimentos realizados desde el paradigma de PRSV empleado en el Estudio 1. El objetivo es estudiar el impacto que tiene emplear las estrategias de tratamiento de datos previas a la estimación y ajuste de modelos (agregado e individual) en las distintas condiciones propuestas de heterogeneidad entre participantes. Chechile (2009) realizó un estudio formal y de simulación en los que comparaba el rendimiento de estas dos estrategias en un grupo de modelos MPT representativos de posibles diferencias estructurales. Sin embargo, dicho estudio no aborda la presencia de heterogeneidad individual como factor. Otros estudios sí se centran en este problema, pero generalmente para evaluar el

rendimiento de una extensión propuesta para abordar condiciones de diferencias continuas (e.g. Klauer, 2010; Matzke et al., 2015; Smith & Batchelder, 2010) o discretas (Klauer, 2006) en el valor de los parámetros bajo el supuesto de que todos los participantes siguen un mismo modelo.

Entre los resultados del Estudio 2, esperamos que en situaciones de homogeneidad el análisis de datos agregados ofrezca estimaciones más eficientes que a partir de datos individuales, debido al sobreajuste al que puede dar lugar el empleo de esta última estrategia. Sin embargo, en condiciones de diferencias discretas debidas a modelos distintos pensamos que el análisis a partir de datos agregados puede enmascarar la presencia de distintos grupos de participantes, ofreciendo estimaciones poco informativas y representativas de los procesos seguidos por los participantes. Estos efectos, en cambio, no aparecerían al realizar la estimación y ajuste a partir de los datos individuales.

1.6.3. Estudio 3. Respecto al rendimiento del modelo jerárquico de rasgo latente en condiciones de heterogeneidad de Matzke et al. (2015)

El modelo de rasgo latente propuesto por Klauer (2010) y recientemente extendido por Matzke et al. (2015), podría superar las limitaciones mencionadas para las estrategias de estimación y ajuste a partir de datos agregados e individuales. Así, a pesar de estar diseñados para abordar diferencias individuales continuas y formalizarlas a partir de distribuciones a priori de carácter unimodal como la normal multivariada pensamos que el modelo podría ser sensible a las distintas fuentes de heterogeneidad individual mencionadas. Consideramos que la descomposición jerárquica de los valores paramétricos que plantea el modelo aporta suficiente flexibilidad como para mostrar sensibilidad a distintas fuentes de variación individual. De hecho, esperamos un comportamiento similar al que ofrece abordar el ajuste a partir de datos individuales, a pesar de dejar sin resolver el problema de clasificación de participantes en subgrupos en caso de existir diferencias discretas.

El objetivo del Estudio 3 es poner a prueba estos supuestos y examinar el rendimiento del modelo en situaciones apenas exploradas, como la presencia de diferencias discretas entre participantes. Para ello simulamos resultados en una tarea de PRSV siguiendo una estructura similar a la del Estudio 2. Posteriormente realizamos la

estimación y ajuste del modelo jerárquico de rasgo latente adaptado para los modelos de árbol sobre CI, analizando su rendimiento en las distintas condiciones de heterogeneidad individual propuestas.

2. Study 1

*On the reality of illusory conjunctions*³

Abstract

The reality of illusory conjunctions in perception has been sometimes questioned, arguing that they can be explained by other mechanisms. Most relevant experiments are based on migrations along the space dimension. But the low rate of illusory conjunctions along space can easily hide them among other types of errors. As migrations over time are a more frequent phenomenon, illusory conjunctions can be disentangled from other errors. We report an experiment in which series of colored letters were presented in several spatial locations, allowing for migrations over both space and time. The distribution of frequencies were fit by several multinomial tree models based on alternative hypothesis about illusory conjunctions and the potential sources of free-floating features. The best-fit model acknowledges that most illusory conjunctions are migrations in the time domain. Migrations in space are probably present, but the rate is very low. Other conjunction errors, as those produced by guessing or miss-categorizations of the to-be-reported feature, are also present in the experiment. The main conclusion is that illusory conjunctions do exist.

2.1. Introduction

In many experimental paradigms studying perception and attention it has been observed that participants sometimes report stimulus compounds that includes features which have been actually presented, but not in the reported combination (e.g., a green R is reported, but a green T and a red R were actually presented). These types of errors

³This paper has been published as Botella, J., Suero, M., & Durán, J. I. (2017). On the Reality of Illusory Conjunctions. *The Journal of General Psychology*, 1-19.

became an important foundation for *Feature Integration Theory* (FIT; Treisman & Gelade, 1980; Treisman & Schmidt, 1982), which used the term *illusory conjunction* (IC) to describe them. Treisman and her colleagues have proposed successive, increasingly sophisticated versions of FIT (Quinlan, 2003), but the idea that the formation of genuine ICs reflects important features of the model has survived. In the first version of the model two sequential stages in the processing of visual stimuli were proposed. In the first stage, simple features of the stimuli are extracted separately (e.g., a shape and a specific color). The information is processed in parallel and at a pre-attentive level, leading to what they called "floating features". Floating features are representations of the features of a given stimulus that are uncoordinated (e.g., a circular shape and the color blue, separately). In the second stage, an integrated visual representation, or percept, is formed by combining or joining the simple features extracted in the previous phase (in the example, a blue circle). In the second stage the percepts are processed in a serial and controlled manner, requiring the involvement of focal attention. Since focal attention is the mechanism that 'glues' floating features, if its action is somehow impeded, it is possible that floating features will be combined incorrectly.

Although in the successive versions of FIT some of its components have evolved, two of its elements have not changed: (a) somehow floating features exist, and (b) they are prone to be combined incorrectly when focal attention is diverted, resulting in real ICs. The existence of ICs is one of the main pillars of the theory because it demonstrates that there is an early stage, prior to the very formation of percepts, when representations of simple features exist.

The FIT was the first model that appeals to the existence of ICs as a source of evidence, as they are a natural by-product of the architecture of the model. Later other models, as that of Botella, Barriopedro, and Suero (2001), also involve the existence of representations of isolated features, although in a different way. As we will explain below, what is important here are not the differences between those models, but the shared assumption that there exist cognitive representations of isolated features. Without ICs part of the empirical evidence that supports FIT and similar models would be lost. In the next section we explain how has been questioned the existence of actual ICs, or rather the soundness of the evidence that support them.

2.1.1. Doubts and debate about the reality of ICs

The term IC, proposed by Treisman and Gelade (1980) can be misleading because instead of being descriptive it implies in itself a proposed way by which the cognitive system produces the type of errors mentioned above, based on the assumptions of the own FIT. At a descriptive level they are *conjunction errors* (two features actually presented, but mismatched in the response) that occur among other types of errors that also appear in the same tasks (e.g., reporting a stimulus compound that includes a feature that was not presented). However, the possibility exists that conjunction errors in the response are not real ‘illusions’ in the sense that they are actually perceived by the observer. They could be, for example, the result of random guessing. This happens when the feature randomly chosen for the response coincides with a feature pertaining to a stimulus in the display. The possibility of a conjunction error produced by mere guessing increases as the set of possible stimulus features in the experimental paradigm becomes small. Conjunction errors can be also the result of confusions in categorizing. For example, when the observer must report the only letter in red and s/he confounds the shape of the red letter. If the red letter is a T but is miss-categorized as an L, and the letter L is also in the display, the observer reports the actually presented letter L. In short, both random guesses and miss-categorized features can yield conjunction errors, but those are not ICs. The label IC implies a very specific assumption about how the cognitive system produces those conjunction errors: the observers perceive something not presented, but formed by a set of features correctly perceived from the stimuli in the display and correctly categorized in a pre-attentive stage. Those representations of isolated features are then miss-combined.

Donk (1999, 2001) has called the existence of IC into question. Importantly, Donk's position is not incompatible with FIT. What she argued was that the evidence of the existence of IC is weak and should be reassessed. Naturally, the credibility of theories as FIT will deteriorate if some of the evidence on which it is based proves to be incorrect. But this would be only a side effect of Donk's arguments. For her the existence of ICs has not been proven, at least if we understand them as the product of an incorrect combination of features correctly identified. There is no doubt that conjunction errors exist, but for Donk these are an experimental artifact that can be explained in more simple ways than actual ICs. Specifically, conjunction errors occur because participants categorize incorrectly one of the features of the target stimulus or because

of mere guessing. Thus, Ivry and Prinzmetal (1991) found that conjunction errors were more probable for more similar features, a result congruent with the idea that at least part of them are produced by incorrect categorizations. However, some authors argue that even after acknowledging those alternative sources of conjunction errors and discounting them, we can still count some conjunction errors as real ICs (Ashby, Prinzmetal, Ivry & Maddox, 1996; Prinzmetal, Diedrichsen & Ivry, 2001). We also believe that not all conjunction errors are the product of guesses or miss-categorizations; part of them are genuine ICs.

We believe that this controversy has occurred by two main reasons. The first one is that it has been addressed by analyzing mainly the ICs that occur in the spatial domain, i.e., between features pertaining to objects presented simultaneously at different spatial locations. That kind of IC is rare and difficult to observe. Instead, ICs in the time domain are much more powerful and frequent phenomena. Lawrence (1971) was the first to describe them, and Botella and others have conducted numerous experiments showing feature migrations in the time domain (e.g., Barriopedro & Botella, 1998; Botella & Eriksen, 1992; Botella, Garcia & Barriopedro, 1992; Botella, Narvaez, Suero, & Juola, 2007); furthermore, they have proposed a model to explain the way they are formed (Botella, Barriopedro & Suero, 2001). The second reason for the controversy is that when analyzing the evidence, it has not been taken into account that in most trials in which the observers report the stimulus correctly, this happens because the focal attention does its job properly. There is some probability that focal attention is properly applied to the target. That probability is a function of factors such as the time the stimuli are displayed, the visual angle, the discriminability of the stimuli, or the SOA in rapid serial visual presentations (RSVP). Only when the focal attention fails other mechanisms able to yield a response are put into play. The frequency of conjunction errors increases as it becomes more difficult to deploy focal attention to the target stimulus. The characteristics of each experimental set determine the probability that focal attention reaches its goal. In addition, these same variables affect the probability for any other mechanism that could be proposed for conjunction errors (responses based on floating features, errors of categorization, pure guesses, and so on). For example, Botella and Eriksen (1991) showed that the rate of conjunction errors increases as the SOA employed in the presentation of a visual series is reduced. So, we don't know the probability of successful focalization of attention for a given experimental condition,

but we can assume that there are a number of hits produced by focal attention. In parallel, other mechanisms can produce hits that are indistinguishable from those produced by focal attention. For example, some hits can be the product of guesses or of ‘fortunate’ responses given from floating features.

2.1.2. Explaining conjunction errors

We consider here two kinds of explanations of how and why the observers make conjunction errors. The first type involves those models who claim that before the system obtains a multidimensional percept, gets isolated representations of the features forming the stimuli. This gives the possibility that these isolated features appear on responses more likely than the features not presented in the trial. Within this type of explanations the most outstanding and influential model is the *FIT* (Treisman & Gelade, 1980), according to which the system works with maps for different dimensions of the stimuli (shape, color, texture, and so on). The focal attention glues them to achieve reportable percepts. However, when attention does not work properly the cognitive system still have the features from the maps. Then, if the experimenter presses the observers to make a response, they will report more probably those features than any (not-presented) feature that is not in the maps. In this group is also the two-stage model of Botella, Barriopedro, and Suero (2001), one specifically designed to explain conjunction errors along the time dimension. It acknowledges two systems. The first one, focal attention, only yields hits. The second one makes a sophisticated guessing taking as a basis the isolated features, and choose among them taking in account their temporal proximity. They come into play sequentially. Only when focal attention does not achieve a satisfactory response the sophisticated guessing mechanism plays any role. The differences between the *FIT* and the two-stage model are not important here. The main point here is that both models involve the presence of representations of isolated features that can play a role when the focal attention fails in generating an integrated percept. Importantly, the existence of conjunction errors via guessing and/or miss-categorizations is compatible with those models, but besides those errors, still exist some conjunction errors that are genuine ICs.

The second type of explanations is the one we consider here as the default hypothesis, as it involves denning the existence of cognitive representations of isolated

features. If the experimental conditions do not allow attention to work properly, then the system simply does not build the multidimensional percept and the observer response should be ‘I don’t know’. All conjunction errors must come from spurious sources, such as guessing or miss-categorizations. Therefore, the observer reports the features at random or with a pattern indistinguishable from random. That is why we will call this the *random hypothesis*. The consequence from this position is that ICs do not really exist. They cannot exist because the system cannot conjoin isolated features if those isolated or free-floating features are not available in the cognitive system. Some methodological artifacts produce the observed conjunction errors. Donk’s position on the non-existence of ICs is congruent with this theoretical position.

The models that acknowledge the existence of representations of isolated features also entail that those features have a greater opportunity of being reported than the non-presented features. On the other side, if we do not acknowledge those representations then the probability of them for being reported is the same as the non-presented features. In the experiment and the models below, we will calculate the probability of reporting both the presented and non-presented features under the random hypothesis. When the proportions of responses of some categories depart from what is expected from the random hypothesis in some predictable way, we will interpret the result as supporting the existence of ICs and, indirectly, of the existence of isolated representations of the presented features.

In short, the main goal of the experiment below is to gather empirical evidence that ICs are real, that not all conjunction errors are the product of methodological artifacts. Our goal is not to provide empirical support specifically to the FIT or the two-stage model. Of course, the existence of ICs is congruent with the models that acknowledge the existence of representations of isolated features, as FIT or the two-stage model. Consequently, if we reach our goal then it is reinforced the credibility of the class of models that acknowledge the existence of representations of isolated features.

Later, having established in the experiment that the ICs exist, we will test some specific hypotheses about the relative importance of several potential sources of ICs. That is, we will explore some intermediate models between the *random hypothesis* (the ICs do not exist) and the general statement of the *FIT* (all features presented have the same status as potential responses). Specifically, we will explore the possibility that the

floating features have some kind of ‘tags’ associated with them that moderate their chances of being selected for the response. In the following experiment, series of stimuli (RSVPs) are presented in several spatial locations. The observer knows the specific locations where the target could appear. We will evaluate the possibility that two of those hypothetical tags are relevant to select the to-be-reported feature: the ‘presented in a location for potential targets’ tag, and the ‘presented at the same location as the target’ tag.

2.1.3. Experiment

The aim of this experiment is to demonstrate that ICs exist, no matter whether in the space or time domain, once taking into account that conjunction errors can be due to other causes. We know from previous literature that it is much easier for ICs to occur in the time domain than in the space domain. In many of the experiments mentioned above the experimental paradigm employed makes ICs rare and difficult to detect because they can only appear as migrations through space. In this experiment ICs in both the spatial and temporal domains are possible. Congruently with previous literature, we expect to find strong evidence of ICs over time and fewer ICs across space.

Colored letters were displayed in RSVPs at several spatial locations; the observers were to report which letter had been presented in a specified color (see Botella et al., 2007). In order to allow richer comparisons we included one experimental condition in which the target could appear in one among two spatial locations (low perceptual load condition, LL), and another experimental condition in which the target could appear in one among four spatial locations (high load condition, HL). Of course, we expect that the overall errors rate will be higher in the HL condition than in the LL condition. However, the relative frequencies of the types of errors must be predictable in some way. If all conjunction errors are due to errors of classification or guessing, then it is possible to calculate the rate of reports of letters (shape features) presented in the several spatial locations, and the rate of non-presented letters. Any departure from those expectations needs an alternative explanation, as for example that ICs indeed exist.

Furthermore, after discussing the experiment we will assess some specific models, as the direction of the departures can tell us something else. The random hypothesis allows making a prediction of the relative frequencies of reports of letters presented in

the same spatial location as the target, presented in other spatial locations than the target, or of non-presented letters. If the relative frequency of errors coming from actually presented stimuli is higher than expected, then the hypothesis that the isolated features are represented in the system is reinforced. Furthermore, we expect to find that the departure is higher for the features presented in the same spatial location as the target (migrations along the time dimension) than for features presented in other spatial locations (migrations along the space dimension). If that happens, then we will conclude that the system takes as more credible for the target the features that share its spatial location.

2.2. Method

2.2.1. Participants

Participants were 24 students from the Autonomous University of Madrid, with normal or corrected vision. The average age of the sample was 18.75 years (SD, 1.81; 19 women, 5 men). They were randomly divided into two experimental groups of equal size: the low perceptual load condition (LL) and the high perceptual load group (HL).

2.2.2. Stimuli and materials

The experiment was programmed in E-prime Version 1.1 (Schneider, Eschman & Zuccolotto, 2002). Several series of 25 colored letters each were randomly generated for each trial. They were either four simultaneous series (LL condition) or six series (HL condition) with the 25 different capitalized letters (all the alphabet except letter M, highly confusable with N). Figure 2.1 shows a diagram with the spatial locations used in the experiment. In the LL condition different series of letters were presented in the outer positions of the central row, while in the HL condition they were presented at the four positions in that row. Before each trial a position was chosen at random in one of the series along the central row in which the letter target would appear. Therefore, the target always appeared in one of two spatial positions in the LL condition and in one of four in the HL condition. Then, it was randomly assigned an ordinal position within the selected series, which could be the 12th, 13th, or 14th positions. The critical set of the target series was formed by five letters: the target letter plus the two letters before and

the two letters after the target. Five different letters were randomly selected for the critical set, and the central letter of the critical set was the target letter. For the positions simultaneous to the critical set in the other series of the same trial five different letters were selected, so that in the time window of the critical set no two identical letters were presented. The only exception was the series of the spatial locations displaced vertically from the fixation point, which were identical; those positions were employed for presenting series of letters in which the observers know that the target cannot be presented, so they don't need to pay attention to them. Therefore, although in the LL condition 4 series were presented, there were only 3 different (in the HL condition 6 series were presented, but only 5 were different). Each of the five letters of the critical set was presented in a different color, randomly assigned (white, yellow, red, green, and blue). The rest of the letters in both the target and the others series, were in different colors, also chosen randomly, bearing in mind the following restrictions: they could not be the color chosen for the target letter, and two consecutive letters in a series could not be in the same color. The background of the screen remained gray throughout the whole session.

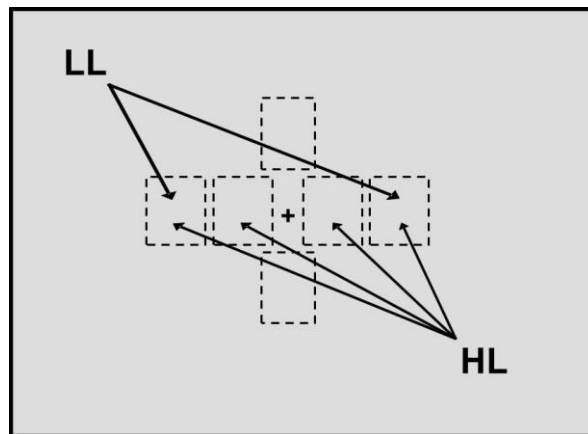


Figure 2.1a. Spatial locations where the series of colored letters were displayed.

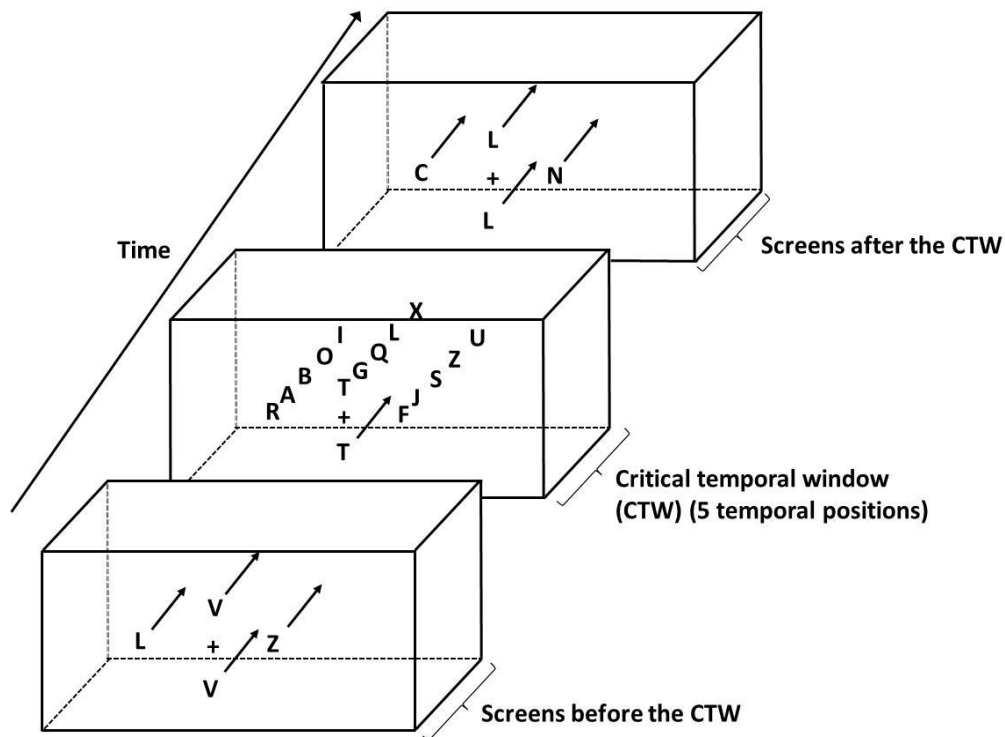


Figure 2.1b. Representation of the positions of the stimuli in space and time in a trial of the LL condition. After a variable number of screens the target letter (the only one in the target color) was presented. The *critical temporal window* (CTW) is defined by the temporal position of the target (the target plus the letters presented two positions before and after it). In the example trial the target is the letter B, presented at the left of the fixation cross. The distribution of the 25 letters employed in the experiment within the CTW is as follows. There are four letters that share the spatial position of the target (R, A, O, I), five letters in the other spatial position where the target could be potentially presented (F, J, S, Z, U), five letters in the two positions vertically displaced, where the observer knows that the target cannot be presented (T, G, Q, L, X). Additionally, there are ten letters that are not presented in the critical temporal window (C, D, E, H, K, N, P, V, W, Y). The arrows within the boxes substitute series of letters, to avoid crowding the figure. The responses of the observers were categorized according to the position in space and time of the letter reported. Then, the expected proportions of responses for an observer that responds completely at random in the LL condition are: 1/25 of hits, 4/25 of intrusions from the same series as the target (SI), 5/25 of intrusions from other locations within the central row, or horizontal intrusions (HI), 5/25 of intrusions from the vertical displaced locations (VI), and 10/25 of non-presented letters, or importations (Imp). For the HL condition, the expected proportions are: 1/25, 4/25, 10/25, 5/25, and 0, respectively.

2.2.3. Procedure

Participants were placed about 40 cm from the screen, from which the letters of the series had a visual angle of 0.46° in height and 0.29° in width. At the beginning of each trial the target color for that trial was set by means of a question as the following: "What letter comes in RED?" with the word RED in red. To begin the trial they pressed the space bar and a cross (+) appeared as fixation point for 500 milliseconds. Then, the series of colored letters appeared on the screen. In the LL condition four series (three different series) were presented with an SOA of 83 msec; the target could appear equiprobably in any of the two locations in the central row. The participants were to report the only letter presented in the target color, by pressing the corresponding key on the keyboard. If they did not know which was the right letter they were told to press the space bar to pass to the next trial. The HL condition was similar to the LL condition, except that four series were presented in the central row instead of two; the target letter could appear randomly in any of those four series. As explained above, in both conditions an additional series was presented above and below the fixation point. The observers knew that the target would never appear in those series.

All participants performed a total of 240 trials. The first 5 were examples employed by the experimenter to explain the task, giving them feedback on their responses. The following 35 trials served as practice to habituate participants to the task. The last 200 trials were the experimental block, although there was no interruption or interval with the practice trials.

2.2.4. Predictions

According to many previous experiments, reporting features from stimuli presented more than two temporal positions away from the target in an RSVP stream is very rare (e.g., Botella & Eriksen, 1991, 1992). As a consequence, we will focus only on the critical temporal window (CTW), defined by the temporal section that includes the screen with the target plus the two screens before and after it (see Figure 2.1b).

We will analyze the data by first assessing the predictions of what we call the *random hypothesis*. Essentially, this hypothesis entails that ICs do not exist. Suppose by now that indeed ICs do not exist and all errors are the result of confusions when

categorizing the features or are produced by mere guessing. As the observer is asked to report the letter in a given color it is possible that in a trial in which the target is a T green s/he confuses the shape of the letter T, for example, with a letter L. That is an error of categorization. It is also possibly a guess, in which the letter L is randomly chosen for the report. How is the response categorized if the observer presses the L key in the keyboard? Given the way the experiment has been designed, the letter L could have been presented in the example trial in any of the locations in the screen (LL and HL condition) or in the pool of non-presented letters (LL condition). Then, the probability that an error is categorized as an intrusion from any position or as an importation depends on the number of letter candidates for those errors. In practice, the classification of the responses is done according to where the letter reported was presented in that trial, within the CTW.

Within the CTW, in each trial of the LL condition there are 4 non-target letters in the spatial location of the target, 5 letters in the other location of the central row, and 5 letters in the locations vertically displaced. Additionally, there are 10 letters not presented within the CTW. The responses were classified as intrusions in the same location as the target (*SI*) when the letter presented was in the critical set of the target series, as horizontal intrusions (*HI*) when the letter was presented within the CTW but in a location of the central row different to that of the target series, as vertical intrusions (*VI*) when the letter was presented within the CTW but in the vertically displaced positions. Finally, the response was coded as an importation (*Imp*) when the letter reported was not presented within the CTW.

Given the number of letters that could be equiprobably classified on each error category under the *random hypothesis*, the probabilities of producing any of them (given that an error is made) are as follows: 4/24, 5/24, 5/24 and 10/24 for the *SI*, *HI*, *VI*, and *Imp* categories in the LL condition; 4/24, 15/24, 5/24, and 0 for the same categories in the HL condition. In short, the prediction from the *random hypothesis* is that the probabilities of the error categories, given that the response is an error, are those in Figure 2.2. It is important to note that we are not saying that the errors of categorizations are committed at random. For example, it is more probable that a red N is confused with an H than with an O because that letter shares more physical characteristics with the target letter. However, the mistaken letter could have appeared

randomly in any position; so, in practice the letter reported (H) would be categorized basically as if it was selected at random as *SI*, *HI*, *VI*, or *Imp*.

To the extent that empirical error rates are not consistent with the *random hypothesis* we could conclude that conjunction errors are not the result of errors of categorizing and/or guessing.

2.3. Results

The main descriptive statistics appear in Table 2.1. As expected, the hit rate was significantly higher in the LL condition than in the HL condition [61.7% vs 46.9%; $t(22) = 3.39$; $p = .003$], as the HL condition is more difficult. Also as expected, the rates of no-responses were not significantly different [10.9% vs 11.0%; $t(22) = 0.035$; $p = .973$]. As a consequence, the overall error rate is necessarily higher for the HL condition. However, we are specifically interested in how the errors are distributed over their several possible sources, especially for the space/time and time migration routes.

Table 2.1. Mean percentages (standard deviations) of each category of responses for the two experimental conditions.

Type of response	Experimental condition	
	LL	HL
<i>Hits</i>	61.7 (8.5)	46.9 (12.5)
<i>SI</i>	21.1 (5.7)	29.0 (7.2)
<i>HI</i>	2.0 (1.1)	10.8 (6.5)
<i>VI</i>	1.2 (0.8)	2.4 (1.5)
<i>Imp</i>	3.1 (2.1)	-
<i>No response</i>	10.9 (7.8)	11.0 (9.8)

Our main analytical strategy in this section consists in to test the predictions of the random hypothesis. Recall that such a hypothesis implies that no matter what stimuli were presented, since the selected feature for the response is taken completely at random among the full set of letters in the task. The following are tests of several specific predictions derived from such hypothesis.

The absolute frequencies of the several types of errors cannot be compared between the conditions, as the overall errors are higher for the HL condition. In order to do proper comparisons we have calculated the rates of each type of errors as a

percentage over the total number of errors. The percentages of errors are as follows. SI: 77.7% (LL) and 69.9% (HL); HI: 7.1% (LL) and 24.7% (HL); VI: 4.4% (LL) and 5.4% (HL); Imp: 10.7% in the LL condition (they were not possible in the HL condition). Only the rates of HI were statistically different [$t(22) = 5.021$; $p < .001$]. This is a trivial result, as the number of letters candidates for a HI was three times higher in the HL than in the LL condition.

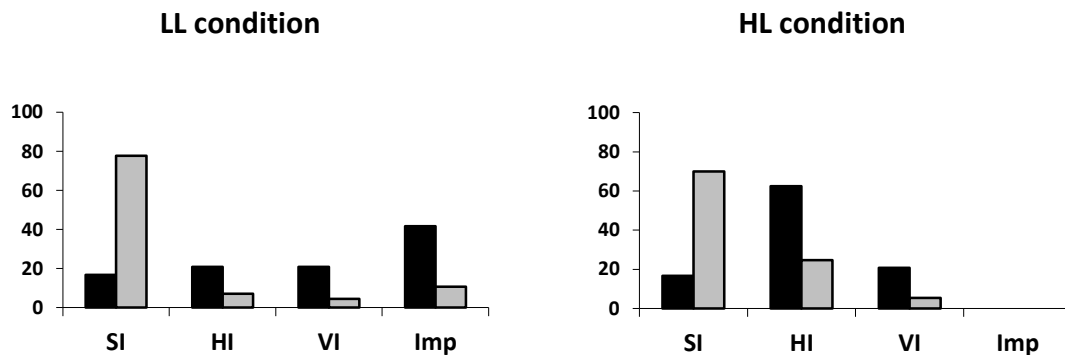


Figure 2.2. Predicted (black bars) and empirical (grey bars) percentages of each type of errors over the total number of errors in the two experimental conditions. Predicted percentages are calculated assuming the *random hypothesis*.

In both experimental conditions, the mean percentage of SI errors greatly exceeded what was expected from the *random hypothesis* (this hypothesis is represented with the black bars in Figure 2.2). This is strong evidence that ICs exist, at least as migrations within the spatial location where the target is presented (migrations in the time domain, or SI errors). The letters presented in the target series, but in a different color than the target color (SI), were much more probably reported than the letters presented in other spatial locations or than the non-presented letters. We tested the hypothesis that the parametric value for the percentage of SI reports is the expected value under the *random hypothesis* ($H_0: \mu = 0.167$). The test yielded a significant departure from those values in both conditions [LL: $t(11) = 21.3$; $p < .001$; HL: $t(11) = 14.0$; $p < .001$].

We can also test whether the percentages of HI and VI responses are congruent with the hypothesis that they are the product of mere guessing. It is predicted from the *random hypothesis* that the *Imp* errors should be twice those of HI and VI errors in the

LL condition. The empirical values were 10.7% of *Imp*, 7.1% of HI and 4.4% of VI. Since under the random hypothesis the expected frequency of IMP errors is twice the expected frequencies of HI and VI errors (see the predictions section), we have tested for no difference between half the empirical frequency of IMP errors and the empirical frequencies of HI and VI errors. That is, we have divided the percentage of *Imp* errors of each observer and have compared those values with the percentages of HI and VI errors with one-sided *t*-tests. The HI are significantly higher than the half values of *Imp* errors [$t(11) = 2.384$; $p = .018$], whereas the VI are not significantly different of the *Imp* errors [$t(11) = -0.849$; $p = .207$]. The second result is congruent with the random hypothesis, but the first one is not. This point to the possibility that migrations in the space domain actually exist, but only from positions attended (positions where the observer expects the target to appear). We need a more powerful tool for analyzing the complex patterns of results involved. In a subsequent section, we will fit some processing models for a finer analysis.

All the above allows rejecting the random hypothesis, as we have found statistically significant departures from its predictions in most of the specific tests performed. We can safely conclude that not all the letters involved in the task have the same probability of being selected for the response. Now we make some comparisons in order to explore the sources of credibility of the features, that is, the properties that moderate the probability that a feature is joined to the target color. For example, the probability of reporting a specific feature could be higher if it has appeared in a spatially close position to the target than if it appeared in a farther one. The credibility of a feature as part of the target could be an inverse function of the spatial proximity between the locations where they were presented. The comparison could only be done in the HL condition. Specifically, we found that the frequency of mis-combinations with letters from the spatial location closest to the location with the target was significantly larger than the other two spatial locations, as revealed by *t*-related tests [comparison with the intermediate location: $t(11) = 5.287$, $p = .0003$; comparison with the farther location: $t(11) = 5.477$, $p = .0002$], while the frequency was statistically equal between the two far locations [$t(11) = 1.265$, $p = .232$].

We also compared the frequency of migrations from the order positions previous to the target with the order positions after the target. Congruently with most previous experiments (Botella, Barriopedro, & Suero, 2001) the frequency of intrusions from

post target positions was significantly larger than from the pre target positions, both for the LL condition [$t(11) = 5.24$, $p = .0003$], and the HL condition [$t(11) = 27.54$, $p < .0001$].

2.4. Discussion

The results show that the distribution of errors between the potential sources is far from random. We can discard the hypothesis that they are completely produced as errors of categorization of the features or as mere guesses. Our main conclusion is that ICs exist, beyond other types of errors that also exist and that can yield conjunction errors.

Our second conclusion is that conjunction errors are much more frequently formed with features sharing the spatial location of the target (migrations in the time dimension, or SI errors) than in any other spatial position (migrations that involve the space dimension, or HI and VI errors). This conclusion is based on several statistical tests of the predictions derived from the random hypothesis, and in the direction of the observed deviations. As shown clearly in Figure 2.2, in both experimental conditions the only type of errors that is statistically more frequent than expected under this hypothesis (grey bar higher than the black bar) are the SI errors (migrations in the time domain). On the contrary, the rates of the other types of errors in both experimental conditions are smaller than expected under the random hypothesis. Within the migrations that involve the space dimension, we found cues that errors are slightly more frequent from spatial locations where the target could appear (in the central row) than from spatial locations where it could not appear (vertically displaced positions). We base this conclusion on that in the LL condition the HI errors are significantly more frequent than expected from the random hypothesis (more than half of the IMP errors), whereas the VI errors are significantly different when making the same comparison. Additionally, the non-negligible percentage of *Imp* responses points that some conjunction errors are produced by guessing or by confusions in categorizing the shape of the target letter.

The high rate of intrusions from the target series indicates that genuine illusory conjunctions exist. Or at least that not all conjunction errors are produced by errors in categorizing the features and/or by random guesses. However, the results suggest also that, except for the SI responses, the rate of ICs from stimuli that do not share the spatial

location of the target are difficult to detect statistically from other types of conjunction errors. There is some processing mechanism that makes conjunction errors involving features actually presented in the same spatial location as the target most likely.

We base our conclusions in a series of statistical tests, comparing mainly the empirical frequencies with those expected from the random hypothesis. However, we need a more complex way to analyze the patterns of results in both conditions. Our design allows, for example, that some ‘fortunate’ guesses give place to hits. In the next section we fit some multinomial tree models that are theoretically motivated and allows a single test that involves the whole distribution of errors.

2.5. Some processing models

We have tested several processing models against the response frequencies observed in the present experiment. The general structure of the models consists of two systems with very different characteristics that work in parallel. On the one hand, a system that operates automatically, fast, and that consumes few resources of processing. On the other hand, a system that works in a controlled, slow and resource-intensive processing (for a similar idea, see Evans, 2008; Kahneman, 2011; Shiffrin & Schneider, 1977). Botella, Barriopedro and Suero (2001) proposed a model with this general structure to explain the formation of illusory conjunctions in the time domain, on which the four models evaluated in this section are based.

In our models it is assumed that the observer can make several attempts to reach a response (Figure 2.3). The observer always tries first to focus attention on the target in order to build a percept that properly reflects the target features, which is successful in the majority of trials. This process is virtually error-free, probably because it involves a mechanism for reviewing the percept. As a consequence, when focal attention is properly deployed, it almost always produces a hit.

It is also assumed that at the end of the processing chain the observer can decide to try a guess (pure guess). The conditional probability that this happens depends on the risk propensity of the participants, given their personal characteristics and the task demands. Both focal attention and guess attempts are common to all the following models. They differ in what they assume happens between the focal attention attempt and the pure guess attempt.

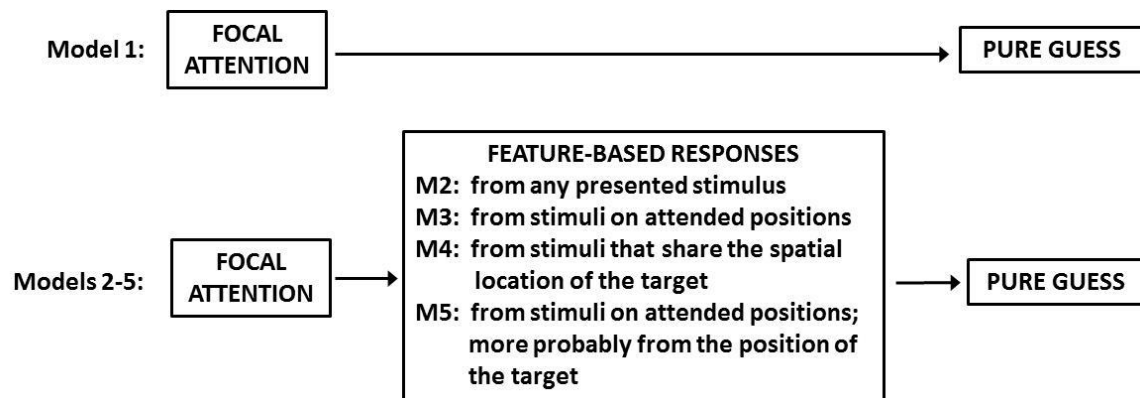


Figure 2.3. Processing models assessed (see text).

When the controlled processing system (which acts by mean of the focal attention) does not get a reportable answer, then before deciding not to answer the observer turns to the outputs of the automatic processing system (models 2-4; see below). Of course, the automatic system is able of getting a potential answer faster. However, as in this experiment the observers were not pressed to make a speeded response we assume that they always give an opportunity to the controlled system to provide an (almost) error-free response. Only when the focal attention fails the observer turns to the automatic system. In the outputs of that system they can find isolated or floating features in which the answer can be based. Botella, Barriopedro and Suero (2001) proposed a mechanism of selection among the floating features available. Finally, if the system does not get the goal, then the observer can decide to respond randomly or choose not to answer.

The M3 and M4 models are a type of intermediate between M1 and M2. Specifically, in M2 the features that can be chosen by the automatic system are those of all the presented stimuli. In contrast, in M3 the set of potential responses is restricted to those that appeared at potential locations for the target. Even more restricted is M4, where the set of potential responses consists only of the features of the stimuli displayed where the target was actually presented. In both M3 and M4 it is assumed that the status of the isolated features is not uniform. The tags associated with them (“potential location for a target” and “spatial location of the target”) moderate the probability that each feature is the one selected for the response.

Let’s see a more detailed description of the four models explored (Figure 2.3).

- *Model 1 (null model)*: ICs do not exist; they are a methodological artefact. All errors are produced by confusions or guessing attempts. As these reports would be in practice classified at random in the error categories of the experiment (*random hypothesis*), their expected frequencies are proportional to the number of possibilities in each category.
- *Model 2 (unconstrained floating model)*: ICs exist and can be formed equiprobably with features presented at any location in the visual field, within the CTW. Of course, this attempt excludes the non-presented features. While the actual ICs must be formed with features actually presented, pure guesses can incorporate both presented and non-presented features.
- *Model 3 (attended locations model)*: ICs exist, but can be formed only by features presented at the attended locations, within the CTW. Attended locations are those where the observer knew that the target could appear: two positions (LL condition) or four positions (HL condition) in the central row. In this model any conjunction errors that include a feature not presented or presented at non-attended positions are the result of erroneous categorizations or guesses.
- *Model 4 (time migrations model)*: ICs exist, but can only come from stimuli that share the spatial position of the target (migrations within the time domain).

Each of the models above makes differential predictions for our experimental task. The parameters of the model are *A* (for Attentional responses), *F* (for **F**eature-based responses), and *g* (for **g**uessing responses). According to M1, when attention fails, the observers sometimes make a guess, whereas sometimes they respond ‘don’t know’ (omissions). As in this model the errors are the result of pure guesses, the probability of reporting a letter from any source is proportional to the number of letters in that source. The conditional probabilities of making a response from each source can be easily calculated for each model. However, we need estimates of the probabilities that focal attention succeeded and that the observer tried a guess. For example, taking only the letters within the CTW in the LL condition there are (besides the target letter, which is coded as a hit when reported) 4 letters in the **SI** source, 5 letters in the **HI** source, 5 letters in the **VI** source, and 10 letters in the **Imp** source. So, the percentages of errors coming from each source should be proportional to those amounts. The only difference for the HL condition is that there is no possibility of **Imp** responses as all 25 letters are presented within the CTW. Furthermore, the percentages of responses from

each source change as the number of letters from each source vary. Now there are 4, 15 and 5 letters in the *SI*, *HI*, and *VI* sources.

M2 assumes that there are free-floating features available after the attentional focus has failed. Those features can come from any position in the display, within the CTW. So, between the focal attention attempt and the pure guess attempt there is another attempt that can only be based in the letters presented in the CTW. Their conditional probabilities are calculated as in M1, but in this attempt *Imp* responses are not possible.

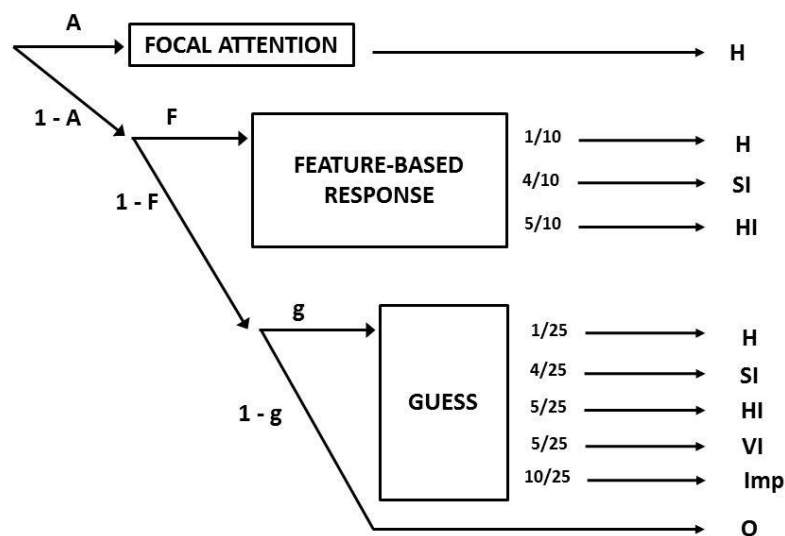


Figure 2.4. Tree diagram of Model 3 for the LL condition. The right end of the diagram represent the types of responses involved (H: hit; SI: intrusion from the same series as the target; HI: intrusion in the horizontal row, that is, from an attended spatial location but different from the location of the target; VI: intrusion in vertical column, that is, from a non-attended spatial location; Imp: importation of a letter not presented within the CTW; O: omission, or no response). The three parameters of the model represent the probabilities of a response yielded by the focal attention (A), by free-floating features coming from the attended locations (F), and from a guess (g). Each branch represent the way (and the probability) of a type of response yielded by a specific route. For example, the probability of a real illusory conjunction from the series in the central row where the target was not presented (HI; **H**orizontal **I**ntrusion) is $(1-A) \cdot F \cdot (5/10)$. Of course, a guess can also be coded as an HI. The overall probability of a given type of response equals the sum of the probabilities of the branches that finish with that type of response. For example, the probability of a hit (H) equals: $A + (1-A) \cdot F \cdot (1/10) + (1-A) \cdot (1-F) \cdot g \cdot (1/25)$.

In M3 and M4 the second attempt is based in free-floating, as in M2. However, the source of the features for such responses varies. In model M3 the only admissible source is the spatial position of one of the potential locations for the target (attended locations; Kawahara, & Miyatani, 2001). In M4 free-floating features are only those presented in the spatial position of the target.

We have fit multinomial tree processing models to the mean frequencies observed (Batchelder & Riefer, 1999; Erdfelder, Auer, Hilbig, Aßfalg, Moshagen, & Nadarevic, 2009). Each model reflects one of the four models just described. To illustrate the structure of the models we show in Figure 2.4 that of M3 for the LL condition.

The estimation and fit of the models was carried out with the package MPTinR (Singmann & Kellen, 2013). Table 2.3 shows the parameter estimates, the degrees of freedom (number of independent amounts minus the number of parameters), and the fit statistic of each model.

Table 2.2. Parameter estimates and fits of the five models assessed.

Model	Low Load condition (LL)				High Load condition (HL)			
	Parameter estimates	<i>df</i>	G^2 (fit)		Parameter estimates	<i>df</i>	G^2 (fit)	
1	<i>A</i>	.61	3	96.7 ($p < .001$)	<i>A</i>	.45	2	113.2 ($p < .001$)
	<i>g</i>	.72			<i>g</i>	.80		
2	<i>A</i>	.60	2	72.2 ($p < .001$)	<i>A</i>	.45	1	113.2 ($p < .001$)
	<i>F</i>	.53			<i>F</i>	.56		
	<i>g</i>	.42			<i>g</i>	.54		
3	<i>A</i>	.59	2	45.6 ($p < .001$)	<i>A</i>	.45	1	97.7 ($p < .001$)
	<i>F</i>	.56			<i>F</i>	.59		
	<i>g</i>	.40			<i>g</i>	.52		
4	<i>A</i>	.56	2	0.4 ($p = .804$)	<i>A</i>	.40	1	0.7 ($p = .393$)
	<i>F</i>	.57			<i>F</i>	.55		
	<i>g</i>	.42			<i>g</i>	.60		

The results show that the empirical frequencies depart significantly from the expected frequencies under models 1-3. Especially important is that the model that explains all conjunction errors as random guesses or by mechanisms that in practice works as such in the present experimental paradigm is incapable of explaining our results. We can safely conclude that in some way ICs exist. Observers are more likely to report a letter presented within the CTW than other features.

On the contrary, model 4 fit the data well in both experimental conditions. The parameters of that model show that the probability of a successful response by focal attention is, as expected, higher for the LL than for the HL condition (.56 vs .40). The conditional probability of a response based on the available free-floating features is almost the same (.57 vs .55). Furthermore, the probability of trying a guess is higher in the HL than in the LL condition (.60 vs .42). Under model 4 the free-floating features available are only those coming from the same series as the target. That is, spatial location plays a fundamental role when trying a response based on isolated features. Probably the mechanism that chooses a feature employs likelihood values, and those values are heavily loaded for features that share the target's spatial location. That is why migrations in space is possible but rare. On the other hand, migrations in time are much more probable. They are a better way to demonstrate that ICs do exist.

2.6. Conclusions

Our main conclusion is that ICs are a real phenomenon. Once other sources of conjunctions errors (confusions and guesses) are taken into account there is a number of responses based on features that were indeed presented, but were not part of the target stimulus. The features selected for the response very frequently share the spatial location with the target. They can also come (probably) from other attended spatial positions. Our best-fitting model is congruent with the so-called two-systems models. One system (the preferred one) is based on focal attention. As it is able of reviewing the display it is virtually error-free. When attention is diverted the other system gives an opportunity to give a response based on sophisticated guessing. It is based on isolated, free-floating features. If that does not work the participant can yet try a guess.

The existence of illusory conjunctions has theoretical consequences for the credibility of models of visual attention and perception. Models of attention should include the possibility that the observers produce illusory conjunctions by means of representations of isolated features. Models as the several versions of the Treisman and her colleagues' FIT or the two-stage model of Botella, Barriopedro, and Suero, are exemplars of the type of models that involve a processing stage in which the system has representations of isolated features. When those representations exist the system has the opportunity of recombine them incorrectly. Any credible model should acknowledge the

evidence involved in the very existence of ICs, as our results demonstrate that they indeed exist.

3. Study 2

Fitting multinomial processing tree models with individual-aggregated data leads to bias when participants follow different models⁴

Abstract

Mathematical models in psychology have often been assessed by fitting aggregated data from a sample of individuals, assuming that all individual data are independent and identically distributed. It has been suggested that fitting models from aggregated data can mask continuous individual differences that would be reflected in high variability among the individual parameters. Individual-by-individual model fitting has been proposed as an alternative data treatment strategy in order to avoid unrealistic unique depictions of heterogeneous samples. We highlight here another heterogeneity source: that within the same sample, subgroups of participants exist that could fit different models. The problem is exemplified with tree models that characterize the formation of illusory conjunctions in attentional tasks. By means of a simulation study, we explore the performance of both - aggregated and individual - data treatment strategies in situations in which: (1) all participants follow a single model with individual variability in parameters, (2) individuals fit different models, and (3) both sources of heterogeneity are present in the sample. We found that fitting models using aggregate data can mask conditions of qualitative heterogeneity, leading to rejection of correct models for a large portion of the sample and to erroneous parametric estimations. Alternatively, the individual fit, though less accurate in parameter estimation, is sensitive to

⁴This study is under review as Durán, J. I., Suero, M. & Botella, J. Fitting multinomial processing tree models with individual-aggregated data leads to bias when participants follow different models.

situations in which individuals fit different models, offering results that faithfully describe the behavior of the whole sample of participants.

3.1. Introduction

The procedure for assessing mathematical models of substantive topics in cognitive psychology has been typically based on estimating parameters and assessing their fit to aggregate data from a sample of individuals (Ashby, Maddox & Lee, 1994). In this process *independent and identically distributed observations (i.i.d.)* are generally assumed to be the case across participants. The variability observed in the sample is attributed to sources external to the model. Such sources include sampling error and random noise due to unknown factors.

In cases where the i.i.d. assumption is met, there would be homogeneity among individuals, so that all their data could be fit by a single model whose parameters take on fixed values.

However, they are conceivable contexts in which the i.i.d. assumption does not hold (Houpt, Yang & Townsend, 2016; Speelman & Muller-Townsend, 2015). This can happen, among other reasons, because:

(a) The entire population fits a single model, but the values of its θ_i parameters vary among the i individuals. This is a situation of *quantitative heterogeneity*, defined as:

$$g(\mathbf{y} | \boldsymbol{\theta}_g) = \int f_{data}(\mathbf{y} | \boldsymbol{\theta}_i) f_{parameters}(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i, \quad (3.1)$$

where \mathbf{y} is the vector of observed results distributed according to g . In statistical terms, the expression is a mixture distribution (Johnson, Kemp & Kotz, 2005) where some or all of the parameters $\boldsymbol{\theta}_i$ involved in f_{data} vary according to a probability density function $f_{parameters}$. Memory deficit models for samples from clinical populations are an example of this situation (Riefer et al., 2002), in which the overdispersion of the parametric values between individuals has been addressed assuming a random effects models.

(b) Individuals fit m different $z = \{1, \dots, z, \dots, m\}$ models without variability in their parameters, resulting in *qualitative heterogeneity*:

$$g(\mathbf{y} | \boldsymbol{\theta}_g) = \sum_{z=1}^m \alpha_z \cdot f_{z,data}(\mathbf{y} | \boldsymbol{\theta}_z), \quad (3.2)$$

where α_z is the proportion of individuals that fit the model z , and $\sum_{z=1}^m \alpha_z = 1$. In statistical terms, the expression refers to a finite mixture distribution (McLachlan & Peel, 2000). Each model z has a different parameter vector θ_z . Those parameters combine in a specific way according to the equations in model z (see section 3.1.2.). As a result, data vector y would also be related to model z equations.

For example, for retention models in Myung, Cavagnaro & Pitt (2017), assuming that four models can coexist in the population, they could be expressed as follows:

For $f_{data, 1}$, y follows the model

$$y = a(t+1)^{-b} \quad (3.3)$$

For $f_{data, 2}$, y follows

$$y = a(t+1)^{-b} + c \quad (3.4)$$

For $f_{data, 3}$

$$y = a(t+1)^{-b} + c + d \cdot \sin(e \cdot t) \quad (3.5)$$

And for $f_{data, 4}$

$$y = ae^{-bt} \quad (3.6)$$

In those expressions, θ_z is determined by the parameters of the models.

It is worth mentioning some examples from psychology that will help contextualize this situation. From a general scientific perspective, one of the main goals of psychology is to develop explanatory theories able to account for individual behavior. However, there are tasks that can be carried out in different ways. For example, given the assumption that the attentional system is common to all "normal" individuals, a theory or general model should be able to accommodate the different strategies or types of behaviors deployed when facing a particular task. These different strategies can be identified with qualitative differences in the models that describe their performances. Such types of differences have been described in several areas within psychology. For example, children can solve addition problems by carrying numbers or by simply counting up from the larger addend. As children grow up, they shift from counting up to carrying numbers (Siegler & Lemaire, 1997). If we propose addition problems to a sample of children at the range of ages when they most frequently move

forward between those strategies, we may end up with a mixture of individuals who use strategies that can be described from qualitatively different models. Similarly, spatial orientation tasks can be undertaken through holistic or analytical strategies (Botella et al., 2009). Many researchers have described strategical changes in different kinds of problems in a variety of areas (Gluck & Fitting, 2003, Gluck, et al., 2001; Lemaire & Siegler, 1995; Peña et al., 2008; Reder, 1987; Reder & Schunn, 1999). In all of them, the situation could be described as a mixture of individuals who perform a given task and their behavior is represented in two or more models whose differences could be characterized as qualitative.

(c) A combination of (a) and (b) or *qualitative and quantitative heterogeneity*:

$$g(\mathbf{y} | \boldsymbol{\theta}_g) = \sum_{z=1}^m \int \alpha_z \cdot f_{z,data}(\mathbf{y} | \boldsymbol{\theta}_i) f_{z,parameters}(\mathbf{y} | \boldsymbol{\theta}_i) d\boldsymbol{\theta}_i, \quad (3.7)$$

where there are participants following m different models. Note that the i.i.d. assumption is a particular case met when $m = 1$, predicting the same parameters vector $\boldsymbol{\theta}$ for all participants.

3.1.1. Group and individual fitting

The data treatment and the subsequent model fit can be approached in two ways: trough individual or group fitting. Individual fitting is carried out for each individual separately, so it offers individual results that can subsequently be summarized with statistics that offer information about central tendency and variability. Alternatively, group fitting consists in aggregating or averaging the individual data to perform a combined single estimation and a unique fitting. The consequences of following either strategy have been studied since the pioneering work by Estes (1956): (a) under conditions of interindividual homogeneity, the group fitting strategy reduces the variability from irrelevant sources for the model. As such, more accurate estimates of the parameter are obtained in comparison to the individual fitting strategy, especially with large samples (Chechile, 2009; Estes & Maddox, 2005). (b) When there is quantitative heterogeneity the group fitting strategy may increase the rate of type I errors, leading to higher frequencies of rejection of valid models (Klauer, 2006). (c) When there is qualitative heterogeneity, the strategy of group fitting does not capture the distinct patterns in task performance that may emerge among different subgroups of

individuals (Cohen et al., 2008), resulting in rejection of models that are correct for subgroups within a population.

This paper examines the consequences of following the individual or group fitting strategies depending on the sources of heterogeneity present in sample (null, quantitative and/or qualitative). In the next section the family of models used in this study, multinomial processing tree models, is sketched. Then, we provide a brief description of the specific phenomenon for which the models employed here have been formulated: illusory conjunctions in time domain.

3.1.2. Multinomial processing tree models

Multinomial processing tree models (MPT) reparametrize probabilities π_j associated with a set of discrete, mutually exclusive and exhaustive $j = \{1, \dots, j, \dots, k\}$ behavioral categories C_j within an specific experimental paradigm (for a more detailed formal description, see Riefer & Batchelder, 1988, and Hu & Batchelder, 1994).

The model structure specifies B_j branches. Each branch r_{bj} is a sequence of S cognitive processes which end in a category C_j , with probability π_j :

$$\pi_j(\boldsymbol{\theta}) = \sum_{b_j=1}^{B_j} P(r_{bj}; \boldsymbol{\theta}), \quad (3.8)$$

where:

$$P(r_{bj}; \boldsymbol{\theta}) = \prod_{s=1}^S \theta_s^{a_{bjs}} \cdot (1 - \theta_s)^{u_{bjs}}, \quad (3.9)$$

$0 \leq \theta_s \leq 1$, and a_{bjs} and u_{bjs} are nonnegative integers which refer to times that a given cognitive process (with probability θ_s) or its absence (with probability $1 - \theta_s$) takes place in a given branch.

The C_j response frequencies follow a multinomial distribution

$$f(\mathbf{x}; n, \boldsymbol{\pi}) = \frac{n!}{x_1! x_2! \dots x_j! \dots x_k!} \pi_1^{x_1} \pi_2^{x_2} \dots \pi_j^{x_j} \dots \pi_k^{x_k}, \quad (3.10)$$

where n refers to the number of trials, x_j to the number of trials in which the result matches the behavioral category j , and π_j to the probability of category j , defined in the vector of parameters $\boldsymbol{\pi}$. The equations involved in each branch r_{bj} of a tree model

specify "processing routes" or combinations of θ_s that would determine each π_j , as expressed in equation (8). In a more general way:

$$\pi_j = h_j(\boldsymbol{\theta}), \quad (3.11)$$

where h_j is the functional relation between θ_s and π_j (see Figure 3.1 for an example). According to equation (11), equation (10) could be re-expressed as:

$$f(\mathbf{x}; n, h(\boldsymbol{\theta})) = n! \prod_{j=1}^k \frac{h(\boldsymbol{\theta})^{x_j}}{x_j!}. \quad (3.12)$$

The sources of heterogeneity highlighted above can be expressed as follows for multinomial distributed data:

(a) Quantitative heterogeneity. In this case, it is assumed that the vector of parameters $\boldsymbol{\theta}_i$ of the MPT model include random variables that follow the $f_{parameters}$ distribution. Considering this, quantitative heterogeneity can be described as

$$g(\mathbf{x}; n, \boldsymbol{\pi}) = \int f(\mathbf{x}; n, h(\boldsymbol{\theta}_i) | \boldsymbol{\theta}_i) f_{parameters}(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i, \quad (3.13)$$

where $f(\mathbf{x}; n, h(\boldsymbol{\theta}_i) | \boldsymbol{\theta}_i)$ is defined in equation (12) for a concrete set of values in $\boldsymbol{\theta}_i$. An example of this formalization can be seen in Smith & Batchelder (2010) beta-MPT, where $f_{parameters}$ is a beta distribution. Other models, as Klauer's (2010) latent-trait MPT, formalize the probit-transformed values of $\boldsymbol{\theta}_i$ with $f_{parameters}$ as a multivariate normal distribution.

(b) Qualitative heterogeneity. The probability of each response category is determined by different MPT models. Thus, in those situations

$$g(\mathbf{x}; n, \boldsymbol{\pi}) = n! \prod_{j=1}^k \frac{\left(\sum_{z=1}^m \alpha_{jz} \cdot h_{jz}(\boldsymbol{\theta}_z) \right)^{x_j}}{x_j!}, \quad (3.14)$$

where $\sum_{j=1}^k \sum_{z=1}^m \alpha_{jz} \cdot h_{jz}(\boldsymbol{\theta}_z) = 1$ and h_{jz} is the functional relation between π_{jz} and $\boldsymbol{\theta}_z$ for model z and response category j . In other words, $h_{jz}(\boldsymbol{\theta})$ defines π for category j following the MPT model z .

(c) Quantitative and qualitative heterogeneity. Since those sources of heterogeneity are not mutually exclusive, both could be present simultaneously, a situation that according to equations (13) and (14) can be expressed as

$$g(\mathbf{x}; n, \boldsymbol{\pi}) = \int n! \prod_{j=1}^k \frac{\left(\sum_{z=1}^m \alpha_{jz} \cdot h_{jz}(\boldsymbol{\theta}_z) \right)^{x_j}}{x_j!} \cdot f_{parameters}(\boldsymbol{\theta}_{iz}) d\boldsymbol{\theta}_{iz}. \quad (3.15)$$

The tree models we use in this study are specifically designed to formally address the results of the rapid serial visual presentation paradigm in the context of illusory conjunctions in time domain (Botella, Barriopedro, & Suero, 2001).

3.1.3. Illusory conjunctions

Illusory conjunctions (ICs) are errors consisting in reporting one stimulus as having properties that do not belong to it, but to a different stimulus presented close to the reported one (close in the visual field or close in time when they are presented in asynchrony). This phenomenon has been employed as evidence to support the feature-integration theory of attention (Treisman & Gelade, 1980). According to this theory, IC are explained as errors in a stage of perceptual processing that requires focal attention to integrate the features extracted from a stimulus to build a single percept. While many authors support this explanation or a similar one (Ashby et al., 1996; Prinzmetal et al., 2001), others argue that the conjunctions can be explained as mere confusions, methodological artifacts, or simply as random responses (Donk, 1999, 2001). Recently Botella, Suero, and Durán (2017) have concluded from an experiment designed to solve the debate that illusory conjunctions actually exist beyond the errors produced by other reasons.

Most evidence examined about this controversy involves the ICs that occur between features of stimuli presented in different spatial locations. However, illusory conjunctions that take place between features of objects presented at different moments but in the same spatial location constitute a more frequent phenomenon. Lawrence (1971) was the first to describe them, and Botella and colleagues have conducted numerous experiments in which they occur (Botella & Eriksen, 1992; Botella et al., 1992; Botella et al., 2007).

Experimental studies on ICs in the time domain commonly make use of the Rapid Serial Visual Presentation (RSVP) paradigm, in which series of typically 15-20 colored letters are presented on each trial. At the beginning of each trial, the letter target for that

trial appears in the screen, in black. The participant began the series by pressing the space bar. Each letter remains on the screen for a short interval (typically, between 70 and 100 msec), being replaced immediately by the next one, and so on until the end of the series. So the rate of presentation is typically about 10-12 items per second; too fast for perceiving all the stimuli. In one of the versions of the task, at the end the observer must report the color in which the target letter (specified before the trial) was displayed. Besides a typical hit rate of about 60%, the errors usually consist in reporting the color of a stimulus presented close in time to the target, with the most frequent error being the report of the color of the stimulus in serial position ‘target+1’ (see Botella & Eriksen, 1991, 1992).

In a more complex version of the task employed by Botella, Suero and Duran (2017), at the end of the trial the observer was to report the only letter presented in the target color. The observer could also make an “I don’t know” response. Furthermore, there were several series simultaneously presented in different locations of the screen. The observer knew in advance that the color target would appear only once in each trial and that it could only appear in some specific series (spatial locations) but not in others. The experimental design allowed as to classify the response in each trial (the letter reported) into five categories unambiguously: (1) hits, when the response matches the target, (2) intrusion of stimuli presented in the same spatial position as the target but at a different moment, (3) intrusion of stimuli presented in a different spatial location than the target, (4) imported stimuli or responses that do not match any letter presented, and (5) omissions or non-responses.

In an experiment of this kind, each participant completes n trials in which she/he gives responses that are distributed among the five categories described above. If X_{ij} is the frequency of responses of participant i in category j , then $\sum_{j=1}^k X_{ij} = n$, where n is assumed equal for all participants.

To model the results of an experiment of this kind, X_{ij} is considered a discrete random variable belonging to a vector of random variables \mathbf{X}_i , where i refers to the participant. This vector follows a multinomial distribution with parameters $(n, \boldsymbol{\pi}_i)$, where n is the number of trials and $\boldsymbol{\pi}_i$ is the vector of probabilities of the response categories for participant i . This vector includes five parameters:

- π_{Hits} : The probability that a successful response occurs.

- π_{IntrusT} : The probability of giving an intrusion from stimuli presented in the same position as the target but at a different moment.
- π_{IntrusS} : The probability of giving an intrusion from stimuli presented in different position as the target.
- π_{Import} : The probability of responding with an imported, not presented, stimulus.
- π_{Omission} : The probability of omitting any response.

The variability in the vector of the values of π , either caused by quantitative, qualitative or both sources of heterogeneity, is acknowledged indicating that vector π_i is linked to participant i . Each value of vector π_i is determined by a series of cognitive processes underlying the response category associated with such value. Therefore, the different values of π_i could be calculated assuming a given set of cognitive processes and the parameters that define them.

The processing routes leading to different response categories can be represented through tree models. In what follows, two tree models are proposed. Each one reflects a different theoretical position about the formation of illusory conjunctions:

Model 1: illusory conjunctions as artifacts. Represents the theoretical position that explains conjunction errors caused by random responses. The model incorporate 2 parameters: F , that quantifies the probability of achieving an adequate focalization of attention, which would lead necessarily to a hit; and g , that refers to random responses following a failure of focal attention, being a process that could lead to any response. Figure 3.1 (upper panel) shows a diagram of the model and the set of equations to obtain the π_j probabilities of each response category depending on the θ_s parameters.

Model 2: illusory conjunctions as real failures of attention. Represents the theoretical position that supports the existence of illusory conjunctions in the time domain, even considering that a subset of observed conjunctions may be the result of random responses. With respect to the previous one, this model includes the parameter R , which quantifies the probability of giving a response based on "floating features", after a failure in focal attention (Botella, et al., 2001). Those "floating features" are available in order to give a response, but are not fully reliable information. Because they are not integrated into a single percept, features originated in the target or in the stimuli presented in the same spatial position as the target but in a different moment can be identified as target features. Therefore, R represents a process that can lead to hits or to intrusions only of stimuli presented before or after the target in its same position

(IntrusT). Figure 3.1 (lower panel) shows the model equations to obtain the probabilities of each response category depending on the parameters values.

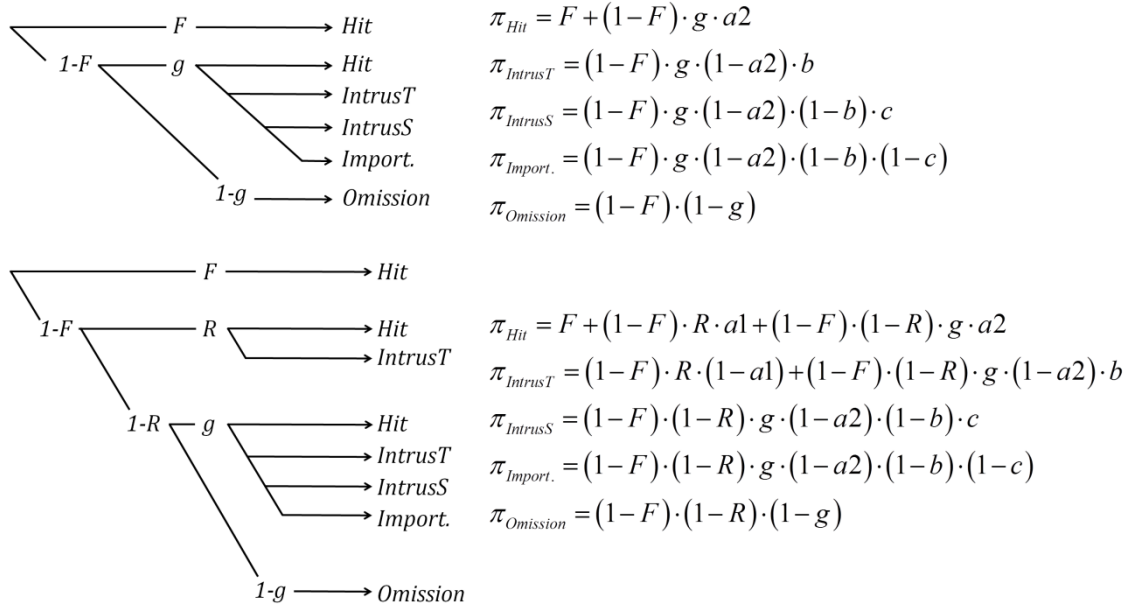


Figure 3.1. Outline and equations of Model 1 (upper panel) and model 2 (lower panel)

Both models depict parameters $a1$, $a2$, b , c , which do not refer to any cognitive process, but establish the probability of choosing a response from the ones belonging to a same response category (see the Method section).

In this study it is assumed that those two models can coexist. From the perspective of the zoom-lens model (Castiello & Umiltà, 1990; Eriksen & St. James, 1986; Eriksen & Yeh, 1985; LaBerge et al., 1997), it is conceivable that some individuals use a narrow attentional focus that shifts to the spatial location where the defining feature of the target is detected. This maximizes the probability of an effective focalization, but it reduces the availability of any floating feature in the system to draw upon when focalization fails. In contrast, other individuals could use a broad focus that would reduce the probability of an effective focalization of attention on the target, but would provide floating features with which to attempt a sophisticated guess (Botella et al., 2001).

This means that in a single sample of participants there can be three possible scenarios: that there are only participants that use a narrow focus, fitting Model 1; only

participants that use a broad focus, fitting Model 2; or there are some participants that fit Model 1 and some that fit Model 2. Accordingly, the first two situations do not involve qualitative heterogeneity, whereas the latter entails qualitative heterogeneity between participants.

In addition, there could be variability in the parametric values among participants that share the same model. In this case, parameters F , R and g are random variables following an unknown distribution. As previously mentioned, this would be a case of quantitative heterogeneity.

Although infinite tree models could be potentially evaluated, those described here are representative of a problem that could appear in real experimental settings when fitting tree models (see Botella, Suero, & Durán, 2017). IC models 1 and 2 define response category frequencies both as a result of unique and multiple tree paths with different structural complexity, yielding a feasible framework to compare group and individual fitting performance in heterogeneity conditions (see Chechile, 2009). In the following, we present a simulation carried out to study the consequences of individual and group strategies when fitting a model depending on the sources of heterogeneity present in the sample (null, only quantitative, only qualitative, or both) in tree models of ICs.

3.2. Method

Response frequency data for each synthetic participant were sampled in a simulation study. To this aim, the values of tree models parameters F , R , and g were specified. Once these values were defined, the values of π from each response category were calculated by equations in Figure 3.1. Given a specific number of trials n , a multinomial distribution with parameters (n, π) was defined for each participant. From this distribution, the frequencies of responses associated with each category were obtained by random generation.

3.2.1. Parameters used in the simulation

In order to obtain the probabilities associated with each response category (π_j) depending on each tree model, parameters F , R and g were set according to

representative values given previous results with the RSVP paradigm and tree models group-estimation from those data (Botella, Barriopedro & Suero, 2001; Botella, Suero & Durán, 2017). Those values appear in Table 3.1.

Table 3.1. Parameter values from models 1 and 2

Model	Parameters (θ_s)		
	F	R	g
1	0.4	-	0.6
2	0.4	0.7	0.6

The values of the remaining parameters ($a1$, $a2$, b and c) were set assuming that all possible responses as equally likely to occur when following each corresponding route in the model. From both models, when a participant guesses a response (with probability g), he/she could respond any possible letter, even those not presented as part of the stimulus in the trial. For the guessing route, $a2$, b and c values are set to reflect the probability of each response category given this assumption (e.g. $a2 = 1/25$ because only one among 25 possible letters match the target). From Model 2, floating features can only be generated from stimulus presented in the same spatial location and different moment as the target. In our RSVP framework only 5 stimuli match this definition, so when a response is due to floating features formation (with probability R), the probability of a hit is $a1 = 1/5$. Complementary, IntrusT probability is $(1 - a1) = 4/5$:

$$a1 = \frac{1}{5}; a2 = \frac{1}{25}; b = \frac{4/25}{(1-a2)}; c = \frac{15/25}{(1-a2) \cdot (1-b)}. \quad (3.16)$$

Setting the number of trials, $n = 200$, the Model 1 follows a multinomial distribution with parameters ($\{.414; .058; .216; .072; .240\}$, 200). Moreover, Model 2 follows a multinomial distribution with parameters ($\{.488; .353; .065; .022; .072\}$, 200).

When adding quantitative heterogeneity, the values of the parameters F , R , and g were simulated following a probit-transformed multivariate normal distribution with variances and covariances as fixed values, presented in Table 3.2.

For each simulation subcondition and model (see next section) the frequencies for 3000 studies with samples of 30 participants were obtained.

Table 3.2. Multivariate normal distribution parameters by quantitative heterogeneity subcondition

Parameter variance	Parameter correlation	MPT parameter	Variance-covariance matrix			
			μ	F	R	g
Low (.01)	Negative (-.3)	F	.4	.01		
		R	.7	-.003	.01	
		g	.6	-.003	-.003	.01
	Null (0)	F	.4	.01		
		R	.7	0	.01	
		g	.6	0	0	.01
	Positive (.3)	F	.4	.01		
		R	.7	.003	.01	
		g	.6	.003	.003	.01
High (.05)	Negative (-.3)	F	.4	.05		
		R	.7	-.015	.05	
		g	.6	-.015	-.015	.05
	Null (0)	F	.4	.05		
		R	.7	0	.05	
		g	.6	0	0	.05
	Positive (.3)	F	.4	.05		
		R	.7	.015	.05	
		g	.6	.015	.015	.05

3.2.2. Simulation conditions

Four sets of simulations were carried out according to the sources of heterogeneity described above:

- *Homogeneity.* Two sets of frequencies were generated. In one of them all participants were generated according to Model 1 and in the other one they all were generated according to Model 2. The parameter values of the MPT were fixed on both cases.
- *Quantitative heterogeneity.* Six sets of frequencies were generated combining different parameter variance and correlation magnitudes. Those $2 \times 3 = 6$ subconditions were the result of combining two levels of parameter variance (low and high magnitude) and three correlation levels (negative, null and positive). Thus, instead of specifying a fixed value for the tree model parameters, values for parameters F , R , and g for each simulated participant were extracted from a

probit-transformed multivariate normal distribution whose mean, variance and covariance parameters are shown in Table 3.2.

- *Qualitative heterogeneity.* Six sets of frequencies were generated in which, for each study, participants generated from models 1 and 2 were combined in different proportions. The proportions employed and the number of participants per study corresponding to each model are presented in Table 3.3. We used non-symmetrical proportions according to the results obtained by Durán, Suero & Botella (2015), in which the estimation process was not sensitive to the increase in the proportion of participants assumed to be following Model 1 until it reaches a proportion greater than .5.
- *Quantitative and qualitative heterogeneity.* The same proportions for combining participants as in the previous condition were specified here. Interindividual variability was also added to MPT model parameters as a factor, sampling tree model parameters for each participant from multivariate normal distributions, whose parameters are shown in Table 3.2. As a result, $6 \times 6 = 36$ data sets were generated by combining the factor levels of quantitative and qualitative heterogeneity.

Table 3.3. Qualitative heterogeneity mixture proportions by subcondition

Subcondition	Proportion N_{Model1}/N_{Total}	N_{Model1}	N_{Model2}
1	0.2	6	24
2	0.5	15	15
3	0.6	18	12
4	0.7	21	9
5	0.8	24	6
6	0.9	27	3

Three-thousand replications, each with $N = 30$ participants and $n = 200$ trials per participant were generated for each condition. The simulations were performed using a program written in R (R Core Team, 2015) and designed by the authors.

3.2.3. Data analysis

Parameter estimation and fit of the two models was carried out with the package *MPTinR v1.6.2* (Singmann & Kellen, 2013). The results were computed from simulated frequencies through ‘fit.mpt’ function (R code available in supplementary materials) with the following strategies:

- (a) Individual. The estimation and fit is performed with each individual set of frequencies. Next, averages (3,000) from the individual estimates belonging to each simulated sample are obtained, from which the final statistics for each simulation (sub)condition are calculated.
- (b) Group (aggregate). Frequencies from each category and participant from each simulated study are summed. The estimation and model fit are carried out from the aggregate data for each sample. The statistics are taken from the results of the 3,000 group estimations.
- (c) Group (averages). The averaged frequencies of the individuals belonging to a same study are used in the estimation. Statistics are obtained for the 3,000 sets of results.

3.3. Results

In order to evaluate the discrepancies between the observed and the maximum-likelihood predicted frequencies, the G^2 statistic (Singmann & Kellen, 2013) is employed. Since the G^2 distribution under null-hypothesis (H_0 : model adequately fit data) is known, we calculate the percentage of samples or individuals in each sample in which each model shows an appropriate fit (% of cases with $p > .05$), so the model is maintained. As G^2 is systematically biased towards more complex models, we also computed the FIA fit index, which takes into account the model complexity and penalizes it via structural complexity. In the following paragraphs we present our main results of Model 1 and Model 2 parameter estimations and fits for each simulated condition. Given that relevant differences in outcomes were not observed between aggregate and average strategies, from now on we refer to both as the group strategy.

3.3.1. Homogeneity

Under homogeneity conditions, the parameter estimates and the fit indices do not show major differences between individual and group strategies. When the estimated model is correct, the parameters do not show bias when compared to their actual values, although individual strategy is less efficient than the group one, showing greater dispersion around the population value.

Regardless of the strategy used, Model 1 holds in about the expected percentage of samples given the value of $\alpha = .05$ when the model is correct (% of $p > .05 = 94.75$), and is rejected in all cases when it is not correct (% of $p > .05 = 0$). Model 2 shows a similar behavior when it is correct (% of $p > .05 = 94.1$), but it is retained in a high percentage of samples in which all participants follow Model 1 (% of $p > .05 = 89.58$). This last result changes when employing the FIA index: when all participants follow Model 1, the percentage of samples on which $FIA_{Model2} > FIA_{Model1}$ equals 0. The high percentage of cases in which Model 2 is retained employing G^2 is due to Model 2's greater flexibility: when estimated with data generated from Model 1 it shows values for the parameter R close to zero, approaching Model 1, in which there are no possible illusory conjunctions due to attention failures. On the other hand, Model 1 parameters show a clear positive bias when estimated using data from Model 2: F and, more notably g , accommodate the frequencies corresponding to the categories in the parameter R branches.

3.3.2. Quantitative heterogeneity

Under conditions of quantitative heterogeneity the results concerning the fit indices are similar. However, even though the variances of the estimates carried out individually are greater than those obtained following a group strategy, they are also much closer to the variance values employed in parameter distributions for random generation.

3.3.3. *Qualitative heterogeneity*

Larger differences between the two fitting strategies are found under conditions of qualitative heterogeneity. The results of the models fit are presented in Figure 3.2. When averaged data are used, the fit index G^2 is insensitive to the increase in the proportion of participants generated from Model 1: only when 90% of the participants or more follow Model 1 is the model retained for all samples. Otherwise, the model is rejected for all samples. However, when using an individual strategy the percentage of cases in which Model 1 is retained grows nearly proportionally to the increase in the number of participants that follow such a model.

With the group strategy Model 2 is retained for about 95% of the replications regardless of the mixture proportion condition. The individual strategy produces results in the fitting that are only slightly sensitive to the change in the proportion of participants from Model 2, showing an over-fit corrected to a great extent when using the FIA index (see Figure 3.2). The better performance with the FIA index is due to its penalty based on the number of parameters or structural complexity in the model. Furthermore, in all the conditions of qualitative heterogeneity the mean percentage of retentions of the true model by individuals and samples is close to the expected value given $\alpha = .05$ (% of $p > .05$, grand mean = 97.74). Therefore, when the group fitting strategy is employed under conditions of qualitative heterogeneity, there is both a tendency to accept a model as valid when it is not (Model 2) and to systematically reject a model (Model 1) that is correct for a large portion of the sample. Those fit results are similar for every quantitative heterogeneity condition regardless of correlation and variance magnitude.

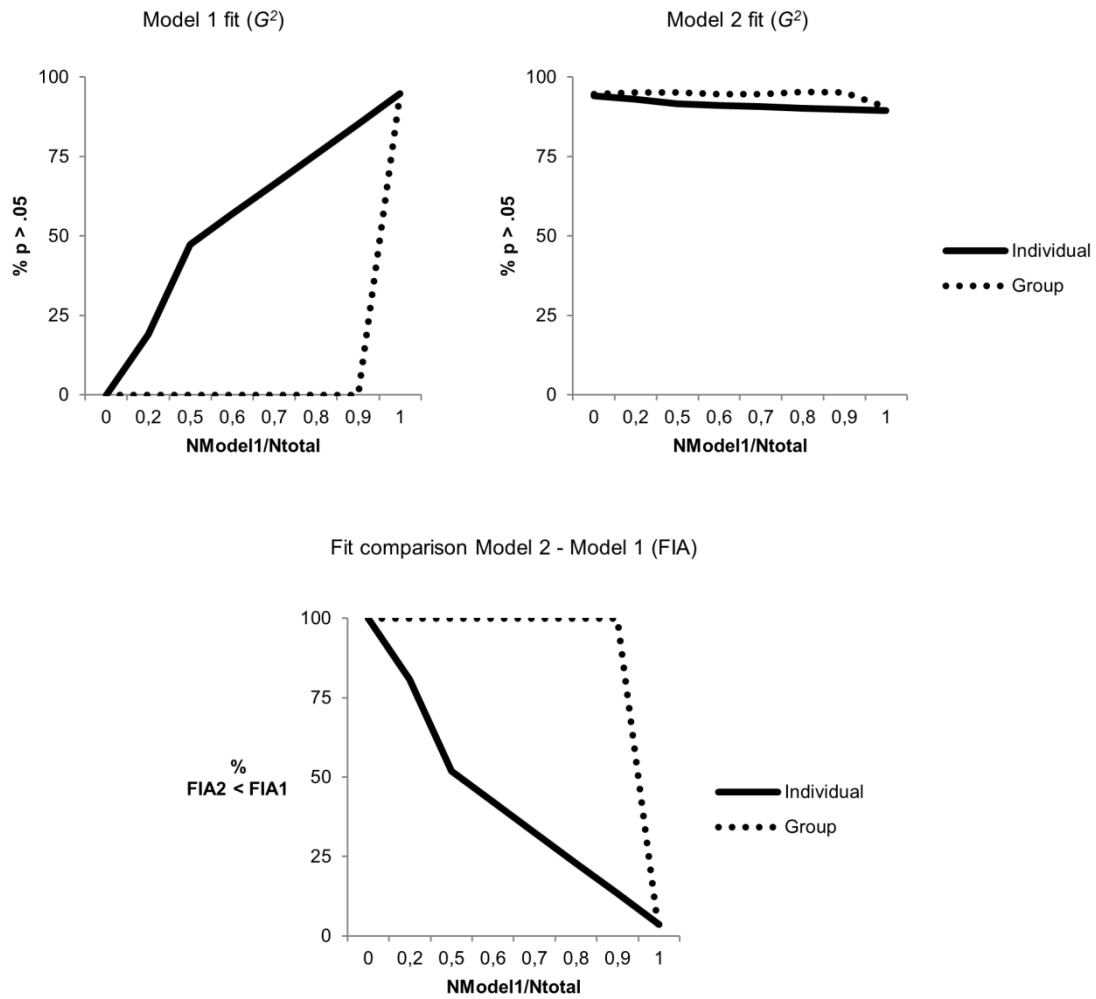


Figure 3.2. Percentage of replications in which the model is retained ($\% p > .05$, upper graphs) and percentage of replications in which $FIA_{Model2} < FIA_{Model1}$ (lower graph) when fitting models with individual and aggregate strategies (by average). Different conditions of qualitative mixture (proportion of participants that follow Model 1) are displayed in x-axis

3.3.4. Effects on the distribution of the estimates

A clear bimodal distribution in Model 1 parameter g values shows up (Figure 3.3A) when estimated following an individual fit strategy with qualitative mixture data. There is also a positive bias in F and g , upon individual estimates of the Model 1 parameters under conditions of qualitative heterogeneity. *IntrusT* is the response category in which the formation of two groups of participants becomes the clearest: one

in which those errors are scarce and due only to random responses and another one in which are frequent because of true attention errors. As the parameter g is the only one that includes the category *intrusT* as an outcome it is the most sensitive parameter to this reality as well. This is a result with high informative value about the presence of qualitative heterogeneity in real contexts, which however does not appear when using a group strategy. When estimating from the averages (Figure 3.3B), the values are distributed around a single value located between both real values, hiding the presence of group mixtures in the sample. This intermediate value varies depending on the proportion of the mixture, and hence with the relative preponderance of each group, approaching more closely to most represented distribution in the sample of participants.

The group strategy effect of estimation convergence into a single modal value is even clearer when fitting Model 2. When representing individual values of the parameter R the estimates cluster into two distributions: one around the value of 0 and one around the value of 0.7 (see Figure 3.3C). The first would correspond to participants who use a narrow focus, and therefore do not show responses based on floating features; the second distribution represent those who employ a broad focus, showing attention errors because of floating feature conjunctions. However, estimation from averages (Figure 3.3D) shows a homogeneous distribution around an intermediate value for R (.35 with a group mixture with proportions 0.5 and 0.5) that does not correctly describe the behavior of any participant and leads to the inference that all of them employ a broad attentional focus. In contrast, very accurate estimates of F and g true values are obtained with the group strategy.

The results hardly vary when adding quantitative heterogeneity in situations of a qualitative mixture of participants, with the exception that a greater dispersion is observed among the estimates.

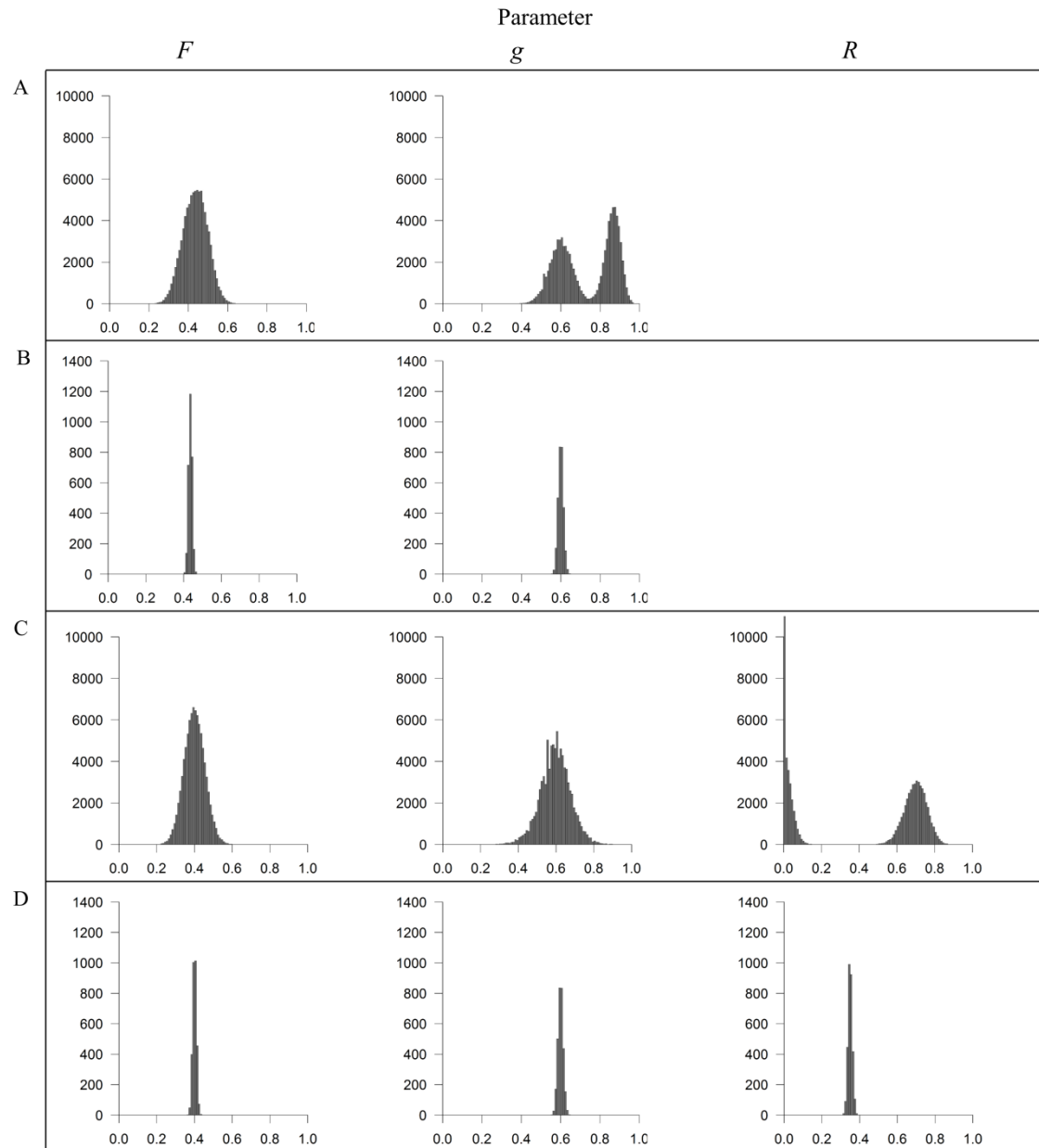


Figure 3.3. Frequency distributions of estimated values for Model 1 parameters from the qualitative mixture data in proportion 0.5 and low quantitative variance ($r = 0$) with individual strategy (A) and group strategy (B). Idem for Model 2 parameters, also for the individual (C) and group (D) strategies.

3.4. Discussion

The results of this simulation study show that under homogeneity among individuals conditions it is preferable to employ an aggregate data strategy in order to obtain a more efficient estimation of the model parameters. This result converges with those from previous studies (e.g., Chechile, 2009; Estes & Maddox, 2005), on which the aggregate or the individual strategy are not advised as ‘absolute’ bests for any situation, but relative to particular research settings. In real situations the population features from which samples are extracted and used for model fitting are unknown. So there is a risk when aggregating heterogeneous data that should actually be considered as samples from different populations (individuals’ behaviors that are generated under different models). As Estes (1956) early pointed out, averaging could substantially modify the results, leading to misinterpretation, especially when employing non-linear models, such as the multinomial processing tree family.

In this paper we consider one among multiple possible situations of qualitative heterogeneity: populations that fit different models of the same family. When this is the case and the individual data are aggregated, ignoring whether the i.i.d. assumption holds, there is a risk of rejecting models that correctly describe the behavior of many of the participants while obtaining parametric estimates describing an unreal average participant. Besides, from the theory testing point of view there is a risk of overrepresentation in the scientific literature of some cognitive processes as typical of the general population when in fact large subpopulations do not engage in such processes at all.

These results make it advisable to follow in general the participant-by-participant fitting strategy in situations where it is suspected that sample participants’ behaviors can be qualitatively heterogeneous. Although there are techniques for detecting overdispersion in the sample (Smith & Batchelder, 2008), these do not distinguish whether heterogeneity is quantitative or qualitative in nature. Given the lack of specificity of this technique, is recommended to take into account that it should yield a positive result in both quantitative and qualitative heterogeneity conditions. Several tests have been developed to detect mixtures of distributions in a sample (Niu, Li & Zhang, 2011). There is also specific software (Benaglia et al., 2009) for detecting mixtures of multinomial distributions in data.

For all these reasons, it is advisable to examine whether there is overdispersion in the sample before estimating and fitting models from aggregate data. If negative results are obtained (i.e. homogeneous data), it would be appropriate to proceed with the usual estimation and fitting procedures from averaged data. However, when overdispersion is detected with any of the techniques mentioned above, it is advisable to use the individual fitting strategy and, only if clearly uniform results are found, proceed to fit models with aggregate data adding a component of random effects to the MPT model parameters (Smith & Batchelder, 2010).

In the present work, we provided classic MPT model results with different strategies to analyze the data in the presence of different sources of individual heterogeneity. It remains for future studies to check the results provided by the recently proposed hierarchical participant heterogeneity models (Klauer, 2010; Matzke et al., 2015) whose aim, among others, is to overcome the limitations of individual and group data treatment strategies. Since it has not been studied, it would be especially interesting to evaluate the performance of those models in qualitative heterogeneity conditions. As hierarchical models take advantage of full individual data, we suspect that the results might be similar to those obtained with the individual data treatment strategy. However, the implementation of hierarchical models is not problem-free, particularly by their complexity and need of further developments of appropriate model checks.

4. Estudio 3

Estimación bayesiana de modelos MPT de rasgo latente con grupos de participantes que se ajustan a modelos distintos⁵

Resumen

La estimación de modelos multinomiales de procesamiento en árbol se realiza frecuentemente a partir de datos promediados por participantes. Como resultado, las conclusiones sobre procesos cognitivos empleados en una tarea experimental se refieren a un único “participante promedio” que puede enmascarar importantes diferencias entre participantes. En la medida en que existan diferencias entre participantes, la generalización de esas conclusiones podría ser sesgada y poco representativa. Las propuestas formales recientemente planteadas para abordar este problema tienen en común el supuesto de que todos los participantes siguen el mismo modelo. En este trabajo exploramos el rendimiento de una de estas propuestas (modelos bayesianos de rasgo latente) en una situación apenas explorada pero plausible. Esto es, cuando la fuente de diferencias entre los participantes es la presencia de subgrupos que ponen en juego distintos procesos cognitivos en la resolución de una misma tarea experimental. En esta situación, los datos de estos subgrupos de participantes se ajustarían a distintos modelos. En el estudio de simulación que hemos llevado a cabo el rendimiento de los modelos de rasgo latente es similar al obtenido ajustando modelos de árbol tradicionales a cada conjunto individual de datos por separado, permitiendo detectar diferencias discretas entre grupos de participantes a través de distribuciones posteriores con un perfil multimodal. Sin embargo, esta ventaja no es tan clara cuando en

⁵Este artículo se encuentra en proceso de revisión como Durán, J. I., Suero, M. & Botella, J. Latent trait MPT bayesian estimation with groups of participants that fit different models.

cada subgrupo existen también diferencias continuas en la probabilidad con la que se emplean los distintos procesos. Los modelos de clase latente presentan también uno de los problemas de la estrategia de ajuste a partir de conjuntos individuales de datos: sobreajuste.

4.1. Introducción

Recientemente ha resurgido el interés por abordar las diferencias individuales en la formalización de modelos cognitivos. Los modelos multinomiales de procesamiento en árbol (MPT) no han sido una excepción, apareciendo distintas propuestas dirigidas a abordar la sobredispersión en los datos debida a las diferencias entre participantes. Una de las propuestas más recientes en este sentido es el modelo jerárquico de rasgo latente de Matzke, Dolan, Batchelder & Wagenmakers (2015). Dado que en este estudio evaluaremos su desempeño en presencia de las fuentes de heterogeneidad detalladas en la introducción de esta tesis, comenzaremos por describirlo formalmente.

4.1.1. Modelo MPT jerárquico de rasgo latente

El modelo de rasgo latente propuesto por Klauer (2010) y extendido por Matzke et al. (2015) formaliza la variación individual descomponiéndola en distintas fuentes que se corresponden con tres niveles jerárquicos (ver Figura 4.1). Un primer nivel hace referencia a la estructura formal del modelo MPT común a todos los participantes. El segundo nivel jerárquico formaliza la variación individual en el vector de parámetros θ del modelo MPT general a partir de una distribución multivariada con tantas dimensiones como elementos tenga el vector θ . El último nivel del modelo de rasgo latente describe la variación en los hiperparámetros de la distribución multivariada del segundo nivel, definiendo la estructura de covariación entre ellos. En los siguientes subapartados trataremos cada nivel jerárquico del modelo en mayor detalle.

4.1.1.1. Nivel 1: Modelo MPT

Hace referencia al modelo de árbol definido formalmente en el apartado 1.5. Se trata de un conjunto de ecuaciones que relacionan las probabilidades de las distintas categorías de respuesta π_j con un vector de parámetros θ cuyos valores son comunes a

todos los participantes. En este nivel del modelo la única variabilidad contemplada es la debida al muestreo.

4.1.1.2. Nivel 2: Variación individual del vector de parámetros θ_i

En este nivel se plantea un modelo de distribución continua para definir las diferencias individuales en los valores de los parámetros del modelo de árbol. El modelo de distribución a priori para la variación en los valores de θ_s es una distribución normal multivariada, con la que se asignan valores de probabilidad para toda la recta real. Al ser probabilidades, los elementos de θ se encuentran en el intervalo $[0,1]$, por lo que para adaptarlos a esta distribución se transforman empleando una función de enlace probit

$$\alpha_{is} = \phi^{-1}(\theta_{is}); \alpha_i \sim N(\boldsymbol{\mu}, \mathbf{V}), \quad (4.1)$$

donde α_{is} es el valor transformado de θ_{is} para el sujeto i -ésimo. Tras esta transformación, el vector de parámetros θ_i pasa a ser α_i . Los parámetros α_i se distribuyen de acuerdo a una distribución normal multivariada con vector de medias $\boldsymbol{\mu}$ y matriz de varianzas-covarianzas \mathbf{V} .

Es importante mencionar que el modelo de rasgo latente descompone los valores de α_i en un componente fijo y uno aleatorio. El componente fijo $\boldsymbol{\mu}$, común a todos los individuos, se corresponde con el vector $\boldsymbol{\mu}$ de la distribución normal multivariada. El componente aleatorio δ_{is} comprende la desviación del i -ésimo individuo⁶ respecto al valor grupal μ_s de acuerdo a un modelo de distribución normal con media 0 y matriz de varianzas covarianzas \mathbf{V}

$$\alpha_{is} = \mu_s + \delta_{is}. \quad (4.2)$$

⁶La propuesta de Matzke et al. (2015) emplea también un término de variación entre las distintas tareas o ítems de un paradigma experimental, que aquí omitimos dado que no está entre nuestros objetivos abordar esta fuente de variabilidad

4.1.1.3. Nivel 3: Matriz de varianzas-covarianzas para los parámetros transformados

El último nivel jerárquico plantea un modelo de distribución para la matriz de varianzas covarianzas V , cuyos valores son variables aleatorias que participan en la definición de la distribución normal multivariada planteada en la ecuación (9). En concreto, el modelo de distribución a priori para la matriz V es la distribución *scaled Inverse-Wishart* (SI-Wishart, Gelman & Hill, 2007) con grados de libertad $S + 1$ y un vector de parámetros de escala ξ con S elementos. La matriz V queda definida como

$$V = \text{Diag}(\xi) \cdot T \cdot \text{Diag}(\xi), \quad (4.3)$$

donde $\text{Diag}(\xi)$ es una matriz diagonal con los parámetros de escala y T sigue una distribución *Inverse-Wishart* con $S + 1$ grados de libertad. Dado este modelo de distribución, las desviaciones típicas σ_s de los parámetros α_s pueden obtenerse mediante la expresión

$$\sigma_s = |\xi_s| \cdot \sqrt{T_{ss}} \quad (4.4)$$

y las correlaciones $\rho_{ss'}$, a partir de

$$\rho_{ss'} = \frac{\xi_s \cdot \xi_{s'} \cdot T_{ss'}}{|\xi_s| \cdot \sqrt{T_{ss}} \cdot |\xi_{s'}| \cdot \sqrt{T_{s's'}}}. \quad (4.5)$$

A pesar de la restricción que supone formalizar la heterogeneidad individual empleando distribuciones continuas unimodales como la distribución normal multivariada y de no estar diseñados específicamente para modelar diferencias discretas, pensamos que el modelo de rasgo latente podría dar cuenta de la presencia de subgrupos de participantes que siguen modelos distintos. El objetivo de este trabajo es poner a prueba esta hipótesis y evaluar rendimiento del modelo de rasgo latente en presencia de distintas fuentes de heterogeneidad entre participantes en un contexto plausible y representativo. Para ello realizamos un estudio de simulación en el que estimamos modelos MPT previamente empleados en investigaciones experimentales sobre atención. El propósito original de estos modelos MPT es arrojar luz en la controversia respecto a la existencia del fenómeno de atención conocido como conjunciones ilusorias (CIs).

Nivel		Modelo
Frecuencias de categoría j	$n_{i1}, ..., n_{ij}, ..., n_{iJ}$	$n_{ij} \sim \text{Multinomial}(n_{ij}, \pi_{ij})$
1. Probabilidad de categoría j	$\pi_{i1}, ..., \pi_{ij}, ..., \pi_{iJ}$	$\pi_{ij} = \sum P(r_{bj}, \theta_i)$
2. Parámetros MPT	$\theta_{i1}, ..., \theta_{is}, ..., \theta_{iS}$	$\alpha_{is} = \Phi^{-1}(\theta_{is})$
Parámetros transformados	$\alpha_{i1}, ..., \alpha_{is}, ..., \alpha_{iS}$	$\alpha_i \sim \text{NormalM}(\mu, V)$
3. Matriz de varianzas-covarianzas	$\begin{pmatrix} \sigma_{11}^2 & & & \\ \vdots & \ddots & & \\ \sigma(s,1) & \cdots & \sigma_{ss}^2 & \\ \vdots & \sigma(s,s') & \vdots & \ddots \\ \sigma(S,1) & \cdots & \sigma(S,s) & \cdots & \sigma_{ss} \end{pmatrix}$	$V \sim SI - \text{Wishart}(gl = S + 1, \xi)$

Figura 4.1. Niveles del modelo jerárquico de rasgo latente (columna izquierda) y modelos de distribución para cada uno de ellos (columna derecha)

4.1.2. Modelos MPT para las conjunciones ilusorias

Como se ha mencionado en la introducción, el fenómeno conocido como conjunción ilusoria (CI) se acuñó a partir de ciertas respuestas erróneas obtenidas en experimentos basados en el paradigma de presentación rápida de series visuales (PRSV; Botella & Eriksen, 1992; Botella, García & Barriopedro, 1992 Botella, Narváez, Suero & Juola, 2007). En un experimento de PRSV se presentan conjuntos de series de letras⁷ coloreados en rápida sucesión. Entre las letras presentadas en un mismo ensayo, solo una aparece en un color específico que el participante sabe de antemano que debe identificar. Esta letra es el estímulo *target*. Al final de la presentación, el participante

⁷Aquí trataremos el caso en el que se presentan letras. Sin embargo, en un experimento de PRSV pueden emplearse otras dimensiones estimulares

reporta la letra que entiende que apareció en ese color. Esta respuesta puede o no coincidir con el estímulo *target*, siendo frecuente observar errores de conjunción, consistentes en reportar como *target* un estímulo realmente presentado, pero que no poseía las características de éste. Así, las respuestas que no coinciden con el *target* se consideran errores, aunque se categorizan de forma distinta en función de su localización en la presentación. Los errores pueden ser: temporales (ET) si se reporta un estímulo presentado en la misma localización espacial que el *target* pero en distinto momento, espaciales (ES) si el estímulo reportado se presentó en una localización espacial distinta o importaciones (Imp) si el estímulo reportado no se presentó.

Los procesos propuestos para explicar los errores de conjunción han supuesto una fuente de debate (Botella, Suero & Durán, 2017) en el que las distintas posturas teóricas pueden formalizarse a través de modelos MPT (Botella, Barriopedro & Suero, 2001).

Las posturas respecto a las CI pueden resumirse en dos. Una de ellas niega la existencia de las CI como un fenómeno debido a procesos psicológicos sustantivos, entendiendo los errores de conjunción como producto de procesos irrelevantes como respuestas al azar. En adelante, esta postura se formaliza a través del Modelo 1. La segunda postura, sin descartar que algunos de los errores puedan deberse a respuestas al azar, explica los errores de conjunción como CIs, es decir, como producto de errores atencionales relacionados con la presentación temporal de estímulos. En adelante, esta postura se formaliza a través del Modelo 2.

Aunque la estructura de los modelos 1 y 2 es distinta, comparten dos de los procesos que proponen, representados por los parámetros: F y g . En cambio, el parámetro R aparece únicamente en el Modelo 2. En la Figura 4.2 se presentan diagramas con las rutas o secuencias de procesos cognitivos que propone cada modelo de árbol para las distintas categorías de respuesta. El parámetro F hace referencia a un proceso de focalización adecuada de la atención, lo que permitiría la elaboración de un percepto que represente con fidelidad el estímulo *target*. Por tanto, este proceso da siempre lugar a respuestas que coinciden con el *target* (aciertos). Por su parte, el parámetro g recoge la probabilidad de dar una respuesta al azar, lo que potencialmente puede dar lugar a cualquier categoría de respuesta exceptuando omisiones. En el Modelo 1 el proceso que representa g solo puede tener lugar tras una focalización inadecuada de la atención y antes de optar por no dar ninguna respuesta (omisión).

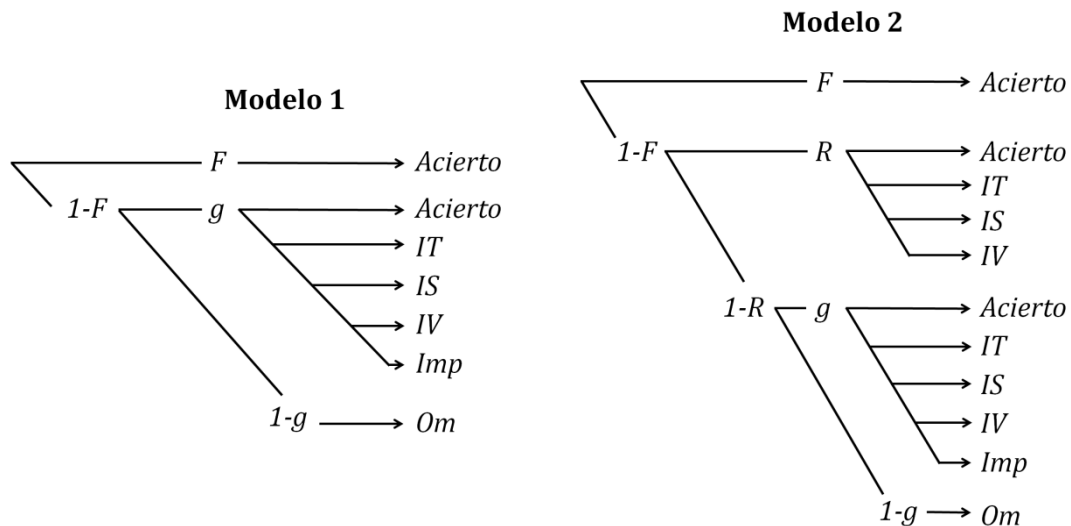


Figura 4.2. Diagramas de las rutas del Modelo 1 (izquierda) y del Modelo 2 (derecha) para las categorías de respuesta. Por ejemplo, desde el modelo 2 las ET se dan como resultado de un fallo en el proceso de focalización ($1-F$), seguido de una respuesta al azar (g). En cambio, el Modelo 2 especifica dos rutas para esta categoría. En una de ellas, se daría un fallo en la focalización ($1-F$) al que sigue una respuesta basada en rasgos aislados (R). En la otra, son fruto de respuestas al azar (g), que únicamente pueden tener lugar una vez que ha fallado la atención focal ($1-F$) y la elaboración de una respuesta a partir de rasgos aislados ($1-R$)

La diferencia principal entre los dos modelos es que el Modelo 2 propone un proceso sustantivo cuya probabilidad queda recogida en el parámetro R . El proceso puede tener lugar tras un fallo en la focalización de la atención y antes de optar por dar una respuesta al azar. Las respuestas resultado de este proceso se basan en la disponibilidad de rasgos aislados no asociados a estímulos concretos, pero procedentes únicamente de estímulos presentados en la misma localización espacial en la que apareció el *target*. Por tanto, este proceso únicamente puede dar lugar a aciertos o a errores temporales (ET). Únicamente los ET especificados en la rama de R se considerarían CIs. Para una explicación más detallada sobre el paradigma de PRSV y el debate respecto de la existencia de las CI que inspira estos modelos ver Botella, Suero y Durán (2017). Las ecuaciones de los modelos de árbol planteados para cada categoría de respuesta en el paradigma PRSV se muestran en la Tabla 4.1.

Al realizar un experimento bajo el paradigma de PRSV, los participantes pueden emplear distintas estrategias para obtener una respuesta válida. Por tanto, al estudiar errores de atención puede haber participantes que empleen, entre otras, estrategias basadas en rasgos aislados (que producirían ET). Del mismo modo, puede haber participantes que no basen sus respuestas en este proceso, centrándose en una focalización de la atención en el estímulo *target*. Al tratarse de una situación plausible, una muestra experimental puede estar compuesta por ambos tipos de participantes “mezclados” en distintas proporciones. Entre ellos, algunos se ajustarían al Modelo 1 y otros al Modelo 2, conformando un caso particular de diferencias discretas debidas a modelos distintos definido en la ecuación (1.4).

Tabla 4.1. Ecuaciones de los modelos MPT empleados en la simulación

Modelo	Categoría de respuesta	Ecuación
1	$\pi_{Acierto} = P(Acierto \theta)$	$F + (1 - F) \cdot g \cdot a2$
	$\pi_{ET} = P(ET \theta)$	$(1 - F) \cdot g \cdot (1 - a2) \cdot b$
	$\pi_{EE} = P(EE \theta)$	$(1 - F) \cdot g \cdot (1 - a2) \cdot (1 - b) \cdot c$
	$\pi_{Imp} = P(Imp \theta)$	$(1 - F) \cdot g \cdot (1 - a2) \cdot (1 - b) \cdot (1 - c)$
	$\pi_{Omisión} = P(Omisión \theta)$	$(1 - F) \cdot (1 - g)$
2	$\pi_{Acierto} = P(Acierto \theta)$	$F + (1 - F) \cdot R \cdot a1 + (1 - F) \cdot (1 - R) \cdot g \cdot a2$
	$\pi_{ET} = P(ET \theta)$	$(1 - F) \cdot R \cdot (1 - a1) + (1 - F) \cdot (1 - R) \cdot g \cdot (1 - a2) \cdot b$
	$\pi_{EE} = P(EE \theta)$	$(1 - F) \cdot (1 - R) \cdot g \cdot (1 - a2) \cdot (1 - b) \cdot c$
	$\pi_{Imp} = P(Imp \theta)$	$(1 - F) \cdot g \cdot (1 - a2) \cdot (1 - b) \cdot (1 - c)$
	$\pi_{Omisión} = P(Omisión \theta)$	$(1 - F) \cdot (1 - R) \cdot (1 - g)$

Consideramos relevante evaluar el rendimiento de propuestas recientes para formalizar diferencias individuales en este tipo de situaciones, ya que la posibilidad de que haya participantes que se ajusten a distintos modelos apenas ha sido explorada en modelos de árbol. Una de las propuestas más recientes es el modelo MPT de rasgo latente de Matzke et al. (2015). Al descomponer los valores de los parámetros en distintos niveles jerárquicos, pensamos que este modelo es lo bastante flexible como para no solo adaptarse a situaciones de diferencias continuas entre participantes, sino como para mostrar sensibilidad a la presencia de diferencias discretas debidas a participantes que siguen distintos modelos. Nuestra hipótesis es que su rendimiento en estas condiciones será similar al obtenido ajustando los modelos MPT a cada conjunto

de frecuencias individual. Sin embargo, el modelo dejaría sin abordar el problema de la clasificación precisa de los participantes en subgrupos.

4.2. Método

Para comprobar estas hipótesis realizamos un estudio de simulación en el que generamos frecuencias correspondientes al paradigma de PRSV en presencia de distintas condiciones de heterogeneidad entre participantes. Se adaptó el modelo de rasgo latente a los modelos de árbol para las CIs, estudiando los resultados de su estimación a partir de las frecuencias generadas bajo las distintas condiciones de heterogeneidad individual.

4.2.1. Valores paramétricos empleados en la simulación

Se simularon frecuencias de respuesta para cada categoría del paradigma de PRSV siguiendo el Modelo 1 y el Modelo 2. Para ello, se fijaron los valores de los parámetros θ_s de cada modelo a valores congruentes a los obtenidos en estudios empíricos en los que se ajustaron a partir de frecuencias grupales. Los valores empleados para los parámetros θ se muestran en la Tabla 4.2, reflejados como μ_s . Una vez fijados los valores y mediante las ecuaciones mostradas en la Tabla 4.1 se pueden calcular las probabilidades π_k de cada categoría. Dadas estas probabilidades y fijando $n = 200$ ensayos por participante simulado, especificamos una distribución multinomial con parámetros (n, π) a partir de la que generar aleatoriamente las frecuencias individuales.

Tabla 4.2. Valores de los parámetros para los modelos MPT en la simulación

Modelo	Parámetros sustantivos (θ_s)			Parámetros fijos		
	$\mu_F(\sigma)$	$\mu_R(\sigma)$	$\mu_g(\sigma)$	$a1=a2$	b	c
1	0.4 (0.1077)	-	0.6 (0.1077)	1/25	4/25	15/25
2	0.4 (0.1077)	0.7 (0.1077)	0.6 (0.1077)		$1 - a2$	$(1 - a2) \cdot (1 - b)$

Siguiendo este procedimiento, en las condiciones de homogeneidad las frecuencias generadas siguiendo el Modelo 1 y del Modelo 2 se obtuvieron a través de distribuciones multinomiales cuyos parámetros aparecen en la Tabla 4.3.

Tabla 4.3. Probabilidades de cada categoría de respuesta para cada modelo MPT en condiciones de homogeneidad

Modelo	$\pi_{Acierto} \theta$	$\pi_{ET} \theta$	$\pi_{EE} \theta$	$\pi_{Imp} \theta$	$\pi_{Omisión} \theta$
1	.414	.058	.216	.072	.240
2	.488	.353	.065	.022	.072

En las condiciones de variabilidad continua (ver siguiente apartado) los valores de θ_s se generaron a partir de una distribución normal truncada para valores alejados en dos desviaciones típicas del valor medio. Los valores empleados en la distribución normal truncada se muestran también en la Tabla 4.2.

4.2.2. Condiciones en la simulación

Las frecuencias se generaron manipulando dos variables de diferencias individuales: el porcentaje de participantes que sigue cada modelo y la presencia o ausencia de variación continua en los parámetros de los modelos. Las condiciones de la simulación son el resultado de combinar los distintos niveles de estas dos variables (ver Tabla 4.4). En concreto, las condiciones del estudio fueron:

- *i.i.d.* Se generaron dos conjuntos de datos: uno desde el Modelo 1 y el otro desde el Modelo 2, en ambos casos con valores fijos para sus parámetros θ_s .
- *Diferencias continuas entre participantes que siguen el mismo modelo.* Las frecuencias se generaron igual que en la condición anterior, aunque agregando variabilidad aleatoria a los valores de los parámetros θ_s de acuerdo a una distribución normal truncada.
- *Diferencias discretas debidas a la mezcla de modelos.* Se generaron seis conjuntos de datos. En cada uno, los participantes que seguían cada modelo (Modelo 1 o Modelo 2) se mezclaron en distintos porcentajes. Los valores porcentuales se mantuvieron fijos para cada estudio simulado dentro del mismo nivel. El porcentaje de casos que seguían el Modelo 1 en los distintos niveles fueron 20, 50, 60, 70, 80 y 90 %.
- *Diferencias discretas debidas a la mezcla de modelos con variabilidad continua en el valor de sus parámetros.* Las frecuencias se generaron del mismo modo que en la condición anterior, aunque agregando variación aleatoria de acuerdo a una distribución normal truncada.

Para cada condición se generaron 1000 estudios con un tamaño muestral $N = 30$ y $n = 200$ ensayos por participante simulado. Las frecuencias simuladas se obtuvieron a partir de un *script* escrito por los autores y ejecutado en R (R Core Team, 2015).

Tabla 4.4. Condiciones en el estudio de simulación como resultado de la combinación de los niveles de las variables porcentaje de participantes que siguen los distintos modelos y presencia de diferencias continuas en el valor de los parámetros

Niveles de mezcla de modelos (% participantes que siguen el Modelo 1 / % Modelo 2)	Niveles de diferencias continuas	
	Efecto fijo	Efectos aleatorios
0 / 100	<i>i.i.d.</i>	<i>Un solo modelo (2) con diferencias continuas en sus parámetros</i>
20 / 80	<i>Diferencias discretas debidas a la mezcla de modelos</i>	<i>Diferencias discretas debidas a la mezcla de modelos con diferencias continuas en sus parámetros</i>
50 / 50		
60 / 40		
70 / 30		
80 / 20		
90 / 10		
100 / 0	<i>i.i.d.</i>	<i>Un solo modelo (1) con diferencias continuas en sus parámetros</i>

4.2.3. Análisis de datos

Adaptamos el modelo de rasgo latente de Matzke et al. (2015) a los modelos 1 y 2. Se aplicó el algoritmo de estimación a cada modelo a partir de las frecuencias individuales correspondientes a cada estudio simulado. Como resultado de este proceso obtuvimos muestras de las distribuciones posteriores de los parámetros en cada uno de los niveles jerárquicos para cada modelo y estudio simulado. En el caso de los parámetros para los que no se definen diferencias entre participantes (por ejemplo, μ_s) se obtuvo una muestra de valores para cada estudio. En cambio, en el caso de los parámetros para los que se proponen variaciones a nivel individual (por ejemplo, α_{is}), se obtuvieron tantas muestras como participantes en cada estudio. Las muestras de las

distribuciones posteriores se obtuvieron con el programa *WinBUGS v1.4.3* (Lunn, Thomas, Best & Spiegelhalter, 2000), y el paquete de R *R2WinBUGS* (Sturtz, Ligges, & Gelman, 2005).

El número de extracciones para muestrear cada distribución posterior se fijó en 10.000 puntos⁸. A partir de estas muestras de la distribución posterior calculamos los valores medios para cada uno de los 30 individuos dentro de cada estudio. A partir de estos valores medios, calculamos los valores medios para cada estudio dentro de una misma condición. Por último, calculamos los valores medios para los distintos estudios dentro de cada condición.

Además, calculamos los valores p bayesianos para cada posterior marginal como la proporción de puntos muestreados que satisfacen la condición de ser superiores o inferiores al valor teórico del parámetro. Valores p de baja magnitud (siguiendo por ejemplo la regla convencional $p < .05$) deben aparecer cuando los valores muestreados que satisfacen la desigualdad son infrecuentes. Tras obtener los valores p bayesianos, calculamos el porcentaje de participantes para los que $p > .05$ en cada estudio simulado, con los que a su vez obtener el porcentaje medio por estudios dentro de cada combinación de niveles de los factores (% medio $p > .05$).

4.3. Resultados

En la Tabla 4.5 se muestran los valores medios obtenidos para cada combinación de niveles de las condiciones. En ella se muestran 5 columnas para cada parámetro del modelo de árbol. De izquierda a derecha, en esta columna se presentan: (1) los resultados medios para la estimación grupal devuelta al rango [0-1], (2) la estimación de las diferencias de cada participante respecto a $\hat{\mu}$, (3) el valor individual estimado para cada individuo devuelta al rango [0-1], (4) el sesgo individual medio respecto al valor teórico del parámetro en escala [0-1] y (5) el porcentaje medio de estudios simulados en los que se obtuvo un valor p bayesiano $> .05$ para una desigualdad respecto a un valor teórico concreto.

⁸Para evitar sesgos debidos a altas correlaciones entre los puntos muestreados de la distribución posterior, sólo una de cada tres iteraciones se retuvo tras descartar las 1000 primeras

Tabla 4.5a. Estadísticos resumen de las distribuciones posteriores estimadas para las condiciones que no incluyen variación continua entre participantes por parámetro del modelo de árbol. Los valores mostrados son los valores medios de los 1000 estudios dentro de cada nivel de mezcla discreta de participantes

Sin variación continua entre participantes

Modelo Ajustado	%M1	Parámetro F (valor teórico; 0.4)					Parámetro R (valor teórico; 0.7)					Parámetro g (valor teórico; 0.6)				
		$\Phi(\hat{\mu})$	$\bar{\delta}_i$	$\hat{\theta}_i$	$(\hat{\theta}_i - \theta)$	% $p > .05$ ($F < .4$)	$\Phi(\hat{\mu})$	$\bar{\delta}_i$	$\hat{\theta}_i$	$(\hat{\theta}_i - \theta)$	% $p > .05$ ($R > .7$)	$\Phi(\hat{\mu})$	$\bar{\delta}_i$	$\hat{\theta}_i$	$(\hat{\theta}_i - \theta)$	% $p > .05$ ($g < .6$)
1	0	0.47	0.0000	0.47	0.07	0.6	-	-	-	-	-	0.86	0.0002	0.86	0.26	0
	20	0.46	0.0009	0.46	0.06	19.7	-	-	-	-	-	0.82	0.0060	0.82	0.22	17.6
	50	0.43	0.0013	0.43	0.03	51	-	-	-	-	-	0.75	0.0070	0.74	0.14	46.4
	60	0.43	0.0010	0.43	0.03	60.4	-	-	-	-	-	0.72	0.0059	0.71	0.12	56.1
	70	0.42	0.0007	0.42	0.02	70.5	-	-	-	-	-	0.69	0.0037	0.69	0.09	65.6
	80	0.41	0.0005	0.41	0.01	80.9	-	-	-	-	-	0.66	0.0029	0.66	0.06	75.4
	90	0.41	0.0002	0.41	0.01	92.3	-	-	-	-	-	0.63	0.0012	0.63	0.03	85.7
	100	0.40	0.0000	0.40	0.01	99.6	-	-	-	-	-	0.60	0.0000	0.60	0.01	99.6
2	0	0.40	0.0000	0.40	0.01	99.6	0.70	0.0000	0.70	0.01	99.4	0.60	0.0004	0.60	0.01	99.6
	20	0.40	0.0000	0.40	0.01	99.6	0.50	0.0040	0.70	0.29	75.6	0.60	0.0003	0.60	0.01	99.6
	50	0.40	0.0000	0.40	0.01	99.6	0.14	-0.1951	0.50	0.50	47.2	0.60	0.0002	0.60	0.01	99.9
	60	0.40	-0.0001	0.40	0.01	99.5	0.06	-0.3638	0.40	0.54	37.8	0.60	0.0002	0.60	0.01	99.9
	70	0.40	-0.0001	0.40	0.01	99.6	0.03	-0.5522	0.30	0.58	28.5	0.60	0.0002	0.60	0.01	99.9
	80	0.40	-0.0001	0.40	0.01	99.7	0.01	-0.7480	0.21	0.62	18.8	0.60	0.0002	0.60	0.01	99.8
	90	0.40	0.0001	0.40	0.01	99.7	0.00	-0.9071	0.11	0.65	9.4	0.60	0.0005	0.60	0.01	99.8
	100	0.40	0.0000	0.40	0.01	99.7	0.00	-0.0204	0.00	0.70	0	0.60	0.0003	0.60	0.01	99.8

$\Phi(\hat{\mu})$: estimación grupal para el parámetro del modelo de árbol correspondiente devuelta a la escala [0,1]

$\bar{\delta}_i$: estimación del parámetro de variabilidad individual

$\hat{\theta}_i$: estimación individual del parámetro del modelo de árbol correspondiente

$(\hat{\theta}_i - \theta)$: sesgo en la estimación individual del parámetro del modelo de árbol correspondiente

Tabla 4.5b. Estadísticos resumen de las distribuciones posteriores estimadas para las condiciones que incluyen variación continua entre participantes

Con variación continua entre participantes

		Parámetro F (valor teórico; 0.4)					Parámetro R (valor teórico; 0.7)					Parámetro g (valor teórico; 0.6)				
Modelo Ajustado	% _{M1}	$\Phi(\hat{\mu})$	$\bar{\delta}_i$	$\hat{\theta}_i$	$(\hat{\theta}_i - \theta)$	% $p > .05$ ($F < .4$)	$\Phi(\hat{\mu})$	$\bar{\delta}_i$	$\hat{\theta}_i$	$(\hat{\theta}_i - \theta)$	% $p > .05$ ($R > .7$)	$\Phi(\hat{\mu})$	$\bar{\delta}_i$	$\hat{\theta}_i$	$(\hat{\theta}_i - \theta)$	% $p > .05$ ($g < .6$)
1	0	0.47	0.0004	0.47	0.08	45.2	-	-	-	-	-	0.88	0.0056	0.87	0.27	0.4
	20	0.45	0.0012	0.45	0.07	50.1	-	-	-	-	-	0.83	0.0122	0.81	0.24	14.1
	50	0.43	0.0010	0.43	0.07	56.7	-	-	-	-	-	0.76	0.0114	0.72	0.20	35.5
	60	0.43	0.0008	0.43	0.07	59	-	-	-	-	-	0.73	0.0093	0.69	0.18	42.8
	70	0.42	0.0002	0.42	0.07	61.9	-	-	-	-	-	0.70	0.0069	0.67	0.16	49.9
	80	0.41	-0.0001	0.41	0.07	64.2	-	-	-	-	-	0.67	0.0055	0.65	0.13	57
	90	0.40	0.0001	0.41	0.07	66.5	-	-	-	-	-	0.64	0.0034	0.63	0.10	64.3
	100	0.40	-0.0010	0.40	0.07	68.7	-	-	-	-	-	0.61	0.0016	0.60	0.07	72.3
2	0	0.39	0.0001	0.40	0.07	72	0.71	0.0034	0.70	0.08	75.7	0.60	0.0028	0.60	0.06	88.2
	20	0.39	0.0001	0.40	0.07	71.8	0.51	0.0069	0.68	0.31	58.5	0.60	0.0032	0.60	0.06	85.6
	50	0.39	-0.0005	0.40	0.07	70.7	0.14	-0.2063	0.50	0.50	36.6	0.60	0.0023	0.60	0.07	81
	60	0.40	-0.0007	0.40	0.07	69.8	0.06	-0.3741	0.40	0.54	29	0.60	0.0028	0.60	0.07	80.1
	70	0.39	-0.0011	0.40	0.07	70.2	0.03	-0.5754	0.30	0.58	21.9	0.60	0.0030	0.60	0.07	78.5
	80	0.40	-0.0013	0.40	0.07	64.2	0.01	-0.7620	0.21	0.62	14.5	0.60	0.0033	0.60	0.07	57
	90	0.40	-0.0015	0.40	0.07	69.5	0.00	-0.9131	0.11	0.65	7.2	0.60	0.0046	0.60	0.07	75.1
	100	0.39	-0.0002	0.40	0.07	68.9	0.00	-0.0227	0.00	0.70	0	0.60	0.0038	0.60	0.07	73

$\Phi(\hat{\mu})$: estimación grupal para el parámetro del modelo de árbol correspondiente devuelta a la escala [0,1]

$\bar{\delta}_i$: estimación del parámetro de variabilidad individual

$\hat{\theta}_i$: estimación individual del parámetro del modelo de árbol correspondiente

$(\hat{\theta}_i - \theta)$: sesgo en la estimación individual del parámetro del modelo de árbol correspondiente

Como era de esperar, en las condiciones de i.i.d. en las que el modelo estimado y el generador de los datos coinciden, la media de la distribución posterior estimada no mostró sesgos. Sin embargo, sí aparece un sesgo positivo en los parámetros del Modelo 1 cuando se estima a partir de datos generados desde el Modelo 2 (0.07 y 0.26 por encima del valor teórico para F y g , respectivamente). Este sesgo positivo aparece de forma sistemática (% medio $p > .05 = 0.6$ y 0 para estimaciones de F y g por debajo del valor teórico). Este sesgo positivo se debe a las altas frecuencias con las que se dan las categorías de respuesta debidas a las rutas de R (aciertos y ET), únicamente presentes en el Modelo 2. El parámetro F y en mayor medida g dan cuenta de las respuestas debidas a las rutas de R . En ausencia de estas últimas, el Modelo 1 únicamente puede explicar las altas frecuencias con las que se dan estas respuestas con estimaciones de probabilidad mayores para F y g .

No aparece sesgo en las distribuciones posteriores de los parámetros F y g del Modelo 2 cuando se estiman con datos generados del Modelo 1 (% medio $p > 0.05 = 99$ para valores de F y g tanto por encima como por debajo del valor teórico). La mayor flexibilidad del Modelo 2 le permite adaptarse a situaciones de ausencia total de CI con distribuciones posteriores de R con valores en torno a 0. Sin embargo, una distribución de estas características no se corresponde con la propuesta sustantiva del modelo: de haber errores debidos a conjunciones en el tiempo y no solo a respuestas al azar, los valores estimados para R deberían ser > 0 .

En condiciones de mezcla de participantes que siguen distintos modelos, la distribución posterior de los parámetros del Modelo 1 muestra un sesgo positivo que decrece a medida que el porcentaje de participantes generados desde este modelo aumenta (Figura 4.3, panel superior derecho). Esta tendencia se da tanto en las estimaciones grupales $\Phi(\hat{\mu})$ como individuales $\hat{\theta}_i$. Los valores del parámetro de variabilidad individual $\bar{\delta}_i$ aumentan a medida que el porcentaje de mezcla de modelos se acerca al valor de máxima heterogeneidad (%_{M1} / %_{M2} = 50/50). Estas consecuencias de la mezcla de participantes que siguen distintos modelos se dan tanto en condiciones de ausencia como de presencia de variabilidad continua en los parámetros del modelo.

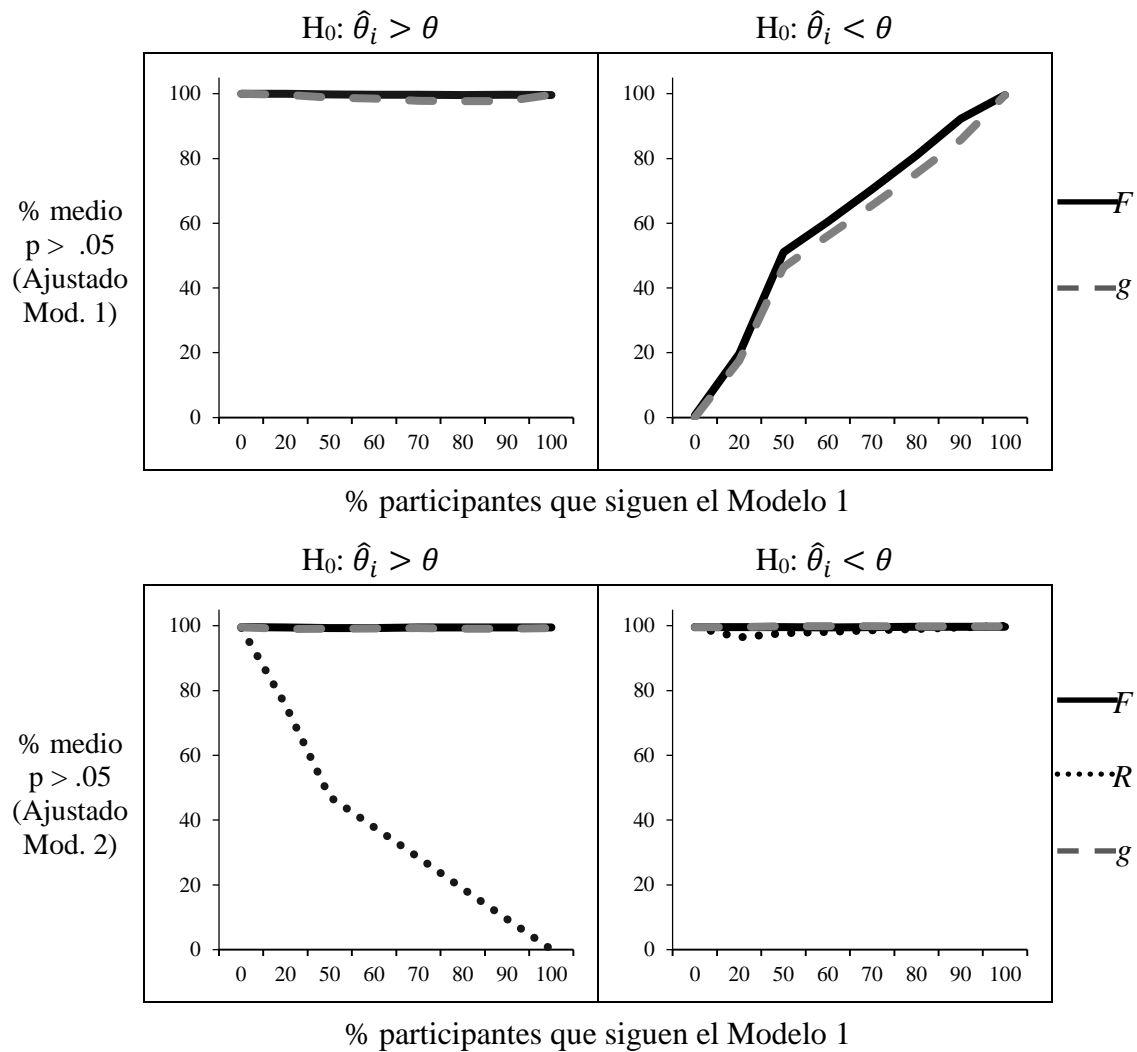


Figura 4.3. Porcentaje medio de participantes por estudio para los que el valor p bayesiano $> .05$. Los valores de $\hat{\theta}_i$ se compararon con los valores teóricos θ_i

En la Figura 4.4 se presentan los histogramas para las muestras de las distribuciones a posteriori obtenidas para los distintos parámetros del modelo de árbol. En concreto, se presentan las distribuciones obtenidas en condiciones de mezcla de participantes que siguen distintos modelos en porcentajes 50/50. Las dos primeras filas de la figura hacen referencia a las estimaciones para el Modelo 1 en ausencia y presencia de variabilidad continua, por lo que los espacios para el parámetro R , al no existir en este modelo, aparecen en blanco. Por su parte, las dos últimas filas de la figura muestran las estimaciones para el Modelo 2. En la distribución posterior de g (Figura 4.4, fila A) se aprecia una acumulación de casos en torno a dos valores modales correspondientes a los dos grupos de participantes que siguen modelos distintos.

Las distribuciones posteriores de los parámetros del Modelo 2 compartidos con el Modelo 1 (F y g) no muestran ningún sesgo o multimodalidad en condiciones de mezcla de participantes que siguen distintos modelos. Al estimar el Modelo 2, solo la distribución posterior del parámetro en el que existen verdaderas diferencias entre participantes (R) se muestra sensible a la mezcla de participantes que se ajustan a distintos modelos (Figura 4.4, fila C). La desviación media respecto al valor grupal ($\bar{\delta}_i$) para R es mayor que la de ningún otro parámetro, y las estimaciones medias individuales $\hat{\theta}_i$ para este parámetro muestran diferencias sistemáticas con las estimaciones grupales $\Phi(\hat{\mu})$ correspondientes. Este efecto se da tanto en ausencia como en presencia de diferencias continuas en los valores de los parámetros (ver Tablas 4.5a y 4.5b). La mezcla de grupos de participantes que siguen distintos modelos también puede apreciarse en la presencia de dos valores modales en la distribución posterior de R : 0 y 0,7. Las estimaciones en torno a cada uno de estos valores se corresponden con los participantes generados desde el Modelo 1 y el Modelo 2, respectivamente. La frecuencia con la que los valores se encuentran cerca de estos valores modales varía con la proporción de participantes que siguen el Modelo 1. Al disminuir la proporción de participantes generados desde el Modelo 1, la frecuencia con la que los puntos muestreados en la distribución posterior de $R > 0.7$ se incrementa, como aparece reflejado en el % medio de $p > .05$ (Figura 4.3, panel inferior izquierdo).

Que uno de los valores modales en la distribución posterior de R se encuentre cerca de 0 puede ser equívoco respecto a la posición más plausible respecto a las CI. A pesar de contrastar el Modelo 2, este resultado supondría una evidencia a favor de la proposición del Modelo 1, al menos para una parte de la población. Además, igual que en los valores en torno al valor modal 0,7, existe variación aleatoria en torno al valor 0 (Figura 4.4, fila C).

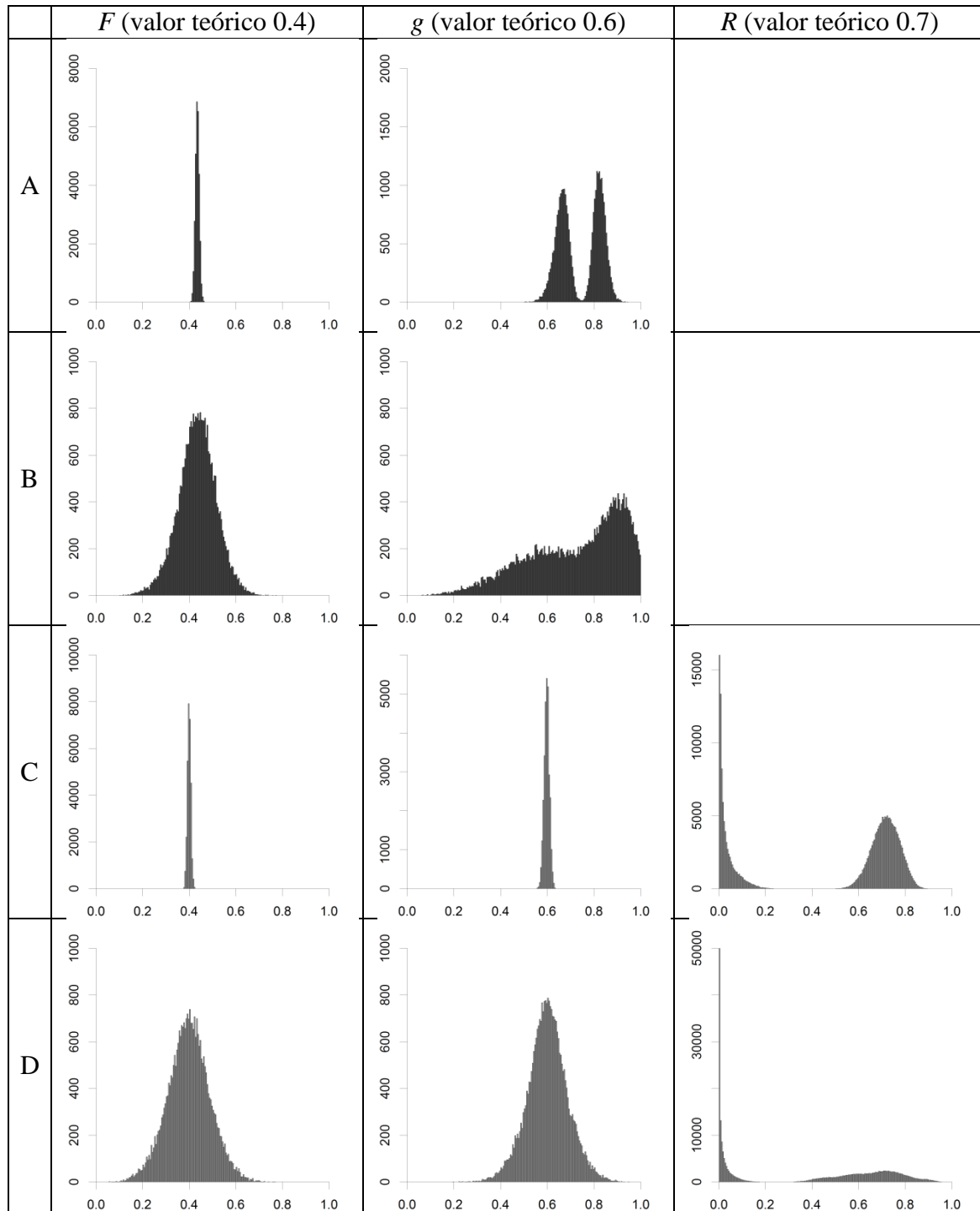


Figura 4.4. Distribuciones posteriores marginales para los parámetros del Modelo 1 en el nivel de mezcla de participantes en porcentajes 50/50 sin variación continua (A) y con variación continua (B). Igualmente para los parámetros del Modelo 2 (C y D)

La magnitud de la variación media ($\bar{\delta}_i$) aumenta en prácticamente todas las condiciones respecto a aquellas en las que no existe esta fuente de variabilidad cuando se estima a partir de datos con diferencias continuas entre participantes. Esta mayor

variabilidad puede apreciarse en los histogramas de las distribuciones posteriores de F y g (Figura 4.4, fila B), en los que la presencia de diferencias discretas debidas a la mezcla de modelos deja de ser claramente identificable a partir de una discontinuidad en el perfil. El perfil multimodal únicamente se mantiene en el histograma de la distribución posterior de R (Figura 4.4, fila D), en la que se sigue apreciando una clara discontinuidad entre los valores 0 y 0.7, muy distantes entre sí. Comparado con el perfil obtenido en ausencia de variación continua, los valores en torno a 0.7 muestran una mayor dispersión, mientras que los valores en torno a 0 se encuentran incluso más concentrados entre ellos.

4.4. Discusión

Los resultados obtenidos muestran un buen rendimiento del modelo jerárquico de rasgo latente cuando la única fuente de heterogeneidad es la mezcla de participantes que se ajustan a distintos modelos. En estas condiciones el modelo es sensible a los cambios en la proporción en la que se da la mezcla de participantes. Así, las distribuciones posteriores permiten inferir la presencia de diferencias discretas a partir de discontinuidades en sus perfiles y la presencia de más de un valor modal. Cada una de estas distribuciones en torno a un valor modal distinto se corresponde con un subgrupo de participantes similares en los datos. No obstante, no podemos descartar que en casos con dos o más subgrupos pero con valores cercanos (e.g., 0,65 y 0,70) no llegue a mostrar una distribución claramente multimodal. En este sentido, el rendimiento del modelo de rasgo latente es similar al obtenido ajustando modelos MPT tradicionales a cada conjunto de datos individual por separado.

El Modelo 1 se podría entender como un caso concreto del Modelo 2 en el que $R = 0$, entendiéndolos como modelos anidados, las condiciones de diferencias discretas empleadas en este estudio de simulación pueden considerarse como de variación entre parámetros de un mismo modelo. No es el caso por dos razones: (1) desde la propuesta del Modelo 1 el valor del parámetro R no es cercano a 0 sino inexistente y (2) al estimar la distribución posterior marginal para el parámetro R aparecen valores *en torno a* 0, y no debería aparecer ninguna variación. Los participantes que siguen el Modelo 1 no recurren a los rasgos flotantes para dar una respuesta con baja probabilidad: no emplean este proceso en absoluto. Las estimaciones del parámetro R que muestran variación

aleatoria continua en torno a 0 para casos en los que el proceso no existe realmente suponen un problema que el modelo de rasgo latente comparte con la estrategia de ajuste de modelos MPT tradicionales a conjuntos de datos individuales: el riesgo de sobreajuste a las particularidades (ruido aleatorio) de cada conjunto de datos.

Por otro lado, la principal ventaja mencionada para el modelo de rasgo latente puede desaparecer cuando dentro de cada grupo que sigue distintos modelos MPT existen diferencias individuales continuas en el valor de sus parámetros. Así, la identificación de subgrupos de participantes que siguen distintos modelos a partir de distribuciones multimodales o discontinuidades se complica o imposibilita en presencia de diferencias continuas dentro de cada subgrupo. Este problema guarda relación con otro que - hasta donde sabemos - aún no tiene solución: no existe una técnica que permita una clasificación precisa de los participantes como pertenecientes a uno de entre múltiples modelos cuando se detectan diferencias discretas entre participantes.

5. Conclusiones generales

Tras la realización de un estudio empírico diseñado para poner a prueba predicciones específicas sobre el origen de los errores de conjunción, encontramos evidencia que apoya la existencia de las conjunciones ilusorias. Esta evidencia se mantiene incluso teniendo en cuenta la posibilidad de que se den otros tipos de errores. Dados los resultados obtenidos estamos en condiciones de concluir no sólo que las CI son un fenómeno real, sino que es más frecuente observarlo en la dimensión temporal que en la espacial. Por su parte, el ajuste de modelos MPT nos permite extraer conclusiones más precisas respecto al modo concreto en que opera el sistema cognitivo en tareas de PRSV. Así, el único modelo que se muestra como plausible a partir de los resultados de nuestro primer estudio plantea dos sistemas de procesamiento. El sistema preferido para dar una respuesta sería lento y operaría bajo un procesamiento controlado (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977): la atención focal. El otro, que forma rasgos flotantes, se dispararía de forma rápida y operaría bajo un procesamiento automático, pero sólo se emplearían las representaciones a las que da lugar para dar una respuesta cuando la atención focal no alcanza su objetivo. El ajuste individual de estos modelos arroja resultados muy similares a los grupales. Sin embargo, existe un pequeño número de participantes para los que incluso este modelo no muestra un buen ajuste. La proporción de casos en los que esto ocurre no fue significativamente diferente de lo esperable por muestreo en caso de que el modelo ajuste correctamente a los datos de todos los participantes. Sin embargo, la posibilidad de que existan diferencias individuales en el modo de abordar una tarea y que por tanto haya participantes cuyos resultados se ajusten a distintos modelos no deja de ser otra explicación plausible, no sólo para el caso de las conjunciones ilusorias sino para otros fenómenos cognitivos.

En el estudio 2 ponemos a prueba esta posibilidad simulando resultados de una tarea de PRSV en presencia de distintas fuentes de variabilidad entre participantes y ajustando dos de los modelos MPT siguiendo distintas estrategias de tratamiento de los datos. Los resultados encontrados en situaciones de homogeneidad son congruentes con estudios anteriores (e.g. Chechile, 2009; Estes & Maddox, 2005). Encontramos que la estrategia grupal rinde estimaciones más eficientes para los valores de los parámetros que la individual. Sin embargo, emplear una estrategia de ajuste grupal sin comprobar si

el supuesto de i.i.d. se cumple supone asumir un riesgo tanto de obtener estimaciones sesgadas para los parámetros como resultados de ajuste poco representativos. En presencia de grupos de participantes que siguen distintos modelos se corre el riesgo de rechazar modelos que describen correctamente el comportamiento de una parte de los participantes. Del mismo modo, existe también el riesgo de mantener modelos como explicaciones generales cuando no describen correctamente el comportamiento de todos los participantes. Dado que la estrategia de tratamiento de los datos más habitual en el ajuste de modelos formales en psicología es la grupal, esto podría estar dando lugar a una sobre-representación de algunos fenómenos en la literatura científica, entendiéndolos como generales o básicos cuando podrían no ser representativos de una parte de la población. Atendiendo a los resultados obtenidos en el estudio 2, emplear una estrategia de ajuste individual es aconsejable cuando se sospecha que en la misma muestra puede haber diferencias entre los participantes en el modo en que abordan la tarea.

A pesar de que las consecuencias de la presencia de diferencias entre participantes en el ajuste de modelos formales en psicología no es una cuestión planteada recientemente (ver Estes, 1956), actualmente está recibiendo bastante atención. Esto se ve reflejado también en el desarrollo de alternativas para los modelos MPT clásicos. Entre las más novedosas se encuentran los modelos jerárquicos de rasgo latente de Klauer (2010) y Matzke et al. (2015), diseñados para dar cuenta de la heterogeneidad continua entre participantes superando las limitaciones de las estrategias de ajuste individual y grupal. En el estudio 3 ponemos a prueba el rendimiento de estos modelos en las distintas situaciones de heterogeneidad descritas, con especial atención a las condiciones de mezcla de grupos de participantes que siguen distintos modelos, dado que es aquella a la que menos atención se ha prestado.

Los resultados obtenidos muestran un buen rendimiento del modelo jerárquico de rasgo latente cuando la única fuente de heterogeneidad es la mezcla de participantes que se ajustan a distintos modelos. En estas condiciones el modelo es sensible a los cambios en las proporciones en las que se presenta la mezcla de participantes. Así, las distribuciones posteriores permiten inferir la presencia de diferencias discretas a partir de discontinuidades en sus perfiles y la presencia de más de un valor modal. No obstante, no podemos descartar que en casos con dos o más subgrupos pero con valores cercanos (e.g., 0,65 y 0,70) no llegue a mostrar una distribución claramente multimodal. En este sentido, el rendimiento del modelo de rasgo latente es similar al obtenido

ajustando por separado modelos MPT tradicionales a cada conjunto de datos individual. Por otro lado, la principal ventaja mencionada para el modelo de rasgo latente puede desaparecer cuando dentro de cada grupo que sigue distintos modelos MPT existen diferencias individuales continuas en el valor de sus parámetros. Así, la identificación de subgrupos de participantes que siguen distintos modelos a partir de distribuciones multimodales o discontinuas se complica o imposibilita en presencia de diferencias continuas dentro de cada subgrupo. Este problema guarda relación con otro que - hasta donde sabemos - aún no tiene solución: no existe una técnica que permita una clasificación precisa de los participantes como pertenecientes a uno de entre múltiples modelos cuando se detectan diferencias discretas entre participantes.

En la operativización de las condiciones de los estudios 2 y 3, si asumimos el Modelo 1 como un caso concreto del Modelo 2 en el que $R = 0$, entendiéndolos como modelos anidados, las condiciones de diferencias discretas podrían considerarse como de variación entre parámetros de un mismo modelo. No es el caso por dos razones: (1) desde la propuesta del Modelo 1 el valor del parámetro R no es cercano a 0 sino inexistente y (2) al estimar la distribución marginal para el parámetro R en ambos estudios aparecen valores *en torno a 0*, y no debería aparecer ninguna variación. Los participantes que siguen el Modelo 1 no recurren a los rasgos flotantes para dar una respuesta con baja probabilidad: simplemente, no emplean este proceso en absoluto. Las estimaciones del parámetro R que muestran variación aleatoria continua en torno a 0 para casos en los que el proceso no existe realmente suponen un problema que el modelo de rasgo latente comparte con la estrategia de ajuste de modelos MPT tradicionales a conjuntos de datos individuales: el riesgo de sobreajuste a las particularidades (ruido aleatorio) de cada conjunto de datos. Por tanto, los resultados obtenidos tanto con la estrategia de ajuste individual como con el modelo de rasgo latente no están exentos de problemas cuando existen grupos de participantes que siguen distintos modelos.

Los resultados de los estudios planteados ponen de manifiesto un problema que aún no tiene una solución formal satisfactoria en psicología y que podría ser, al menos en parte, una de las causas de algunos debates especialmente relevantes en esta disciplina. Entre estos debates figurarían, por ejemplo, los resultados incongruentes obtenidos entre las escuelas que defienden la consistencia del comportamiento de distintos individuos bajo ciertos condicionantes situacionales, por una parte, y aquella que defiende la relevancia de las diferencias entre individuos independientemente del

contexto, por otra (Ross & Nisbett, 2011). De este modo, es posible que existan individuos más sensibles a determinados factores situacionales que otros, lo que daría lugar a dos grupos de participantes con un comportamiento similar dentro del mismo grupo, pero cualitativamente distinto entre ellos al enfrentarse a una situación similar. Es posible que el fenómeno planteado aquí pueda estar también detrás de al menos algunos de los resultados que han dado lugar a la llamada crisis de replicabilidad en psicología (Open Science Collaboration, 2015; Pashler & Wagenmakers, 2012). Precisamente por la tradición de tratar los datos de forma grupal sin comprobar si se cumple el supuesto de i.i.d., y en línea con los resultados obtenidos aquí, en caso de que existan grupos de participantes cualitativamente distintos en un fenómeno bajo estudio es posible que algunos trabajos aporten evidencia de la naturaleza general de un fenómeno mientras otros no encuentren evidencia que lo apoye en absoluto.

5.1. Sugerencias para el ajuste de modelos formales en investigación aplicada

Para evitar en la medida de lo posible los riesgos mencionados en el trabajo con modelos formales, consideramos importante explorar la posibilidad de que existan diferencias entre los participantes en el fenómeno bajo estudio. Para hacer esto, es necesario en primer lugar trabajar con diseños en los que se recojan resultados para un número suficiente de ensayos por participante, lo que permite estudiar el patrón de variabilidad de cada individuo. A partir de este tipo de datos es posible explorar si existen patrones sistemáticos que pudieran repetirse únicamente en subgrupos de individuos. Existen técnicas para detectar si la variabilidad individual es mayor de lo esperable por muestreo (por ejemplo, Smith & Batchelder, 2008), aunque éstas no permiten determinar si las fuentes son de naturaleza continua o discreta, arrojando resultados positivos en ambos casos. Se han desarrollado también pruebas (Niu, Li & Zhang, 2011) e incluso software específico (Benaglia, Chauveau, Hunter & Young, 2009) para detectar mezclas de distribuciones en una muestra.

Los resultados de estos estudios preliminares pueden ser valiosos de cara a escoger la estrategia de ajuste más adecuada para cada caso, ya sea ajustar los modelos a cada conjunto individual de datos o a un único conjunto de datos agregados o promediados por participantes. En función de la evidencia obtenida respecto a la

existencia de heterogeneidad entre individuos también se puede evaluar la conveniencia de aplicar adaptaciones para los modelos a ajustar que permitan dar cuenta de distintas fuentes de heterogeneidad entre individuos. Para el caso de los modelos MPT, entre estas adaptaciones se encontrarían los modelos jerárquicos de clase latente (Klauer, 2006) para diferencias discretas dentro de un mismo modelo, o los de rasgo latente (Klauer, 2010; Matzke et al., 2015), para diferencias continuas. Es importante mencionar que para el caso de los modelos de árbol no existen adaptaciones específicamente dirigidas a formalizar la presencia de participantes que siguen modelos distintos, por lo que el trabajo se centraría en valorar los resultados de la estimación y ajuste individual de todos los modelos formales contemplados.

5.2. Limitaciones y líneas futuras

Es importante tener en cuenta que la forma en la que hemos operativizado en los estudios 2 y 3 las condiciones de diferencias entre participantes debidas a modelos distintos es solo un ejemplo particular, aunque plausible y vinculado a un diseño de investigación real. En concreto, esta operativización comprende modelos de una misma familia (modelos MPT) para respuestas categóricas, que además comparten algunos de los procesos que proponen. Las diferencias cualitativas entre participantes debidas a modelos distintos pueden tomar otras formas. Por ejemplo, cuando el comportamiento de distintos grupos de individuos se ajusta a modelos de distintas familias, algunos de los resultados obtenidos podrían diferir de los expuestos aquí.

Por su parte, el problema planteado puede tener otras repercusiones en modelos diseñados para formalizar datos de distinta naturaleza. Por ello, las garantías de generalización de los resultados obtenidos a, por ejemplo, la aplicación de modelos para formalizar respuestas continuas, como tiempos de reacción, son limitadas. Dejamos estas líneas de desarrollo para futuras investigaciones con las que pretendemos seguir el camino comenzado en esta tesis.

5. General conclusions

The first experimental study is designed to test main explanations and predictions about illusory conjunctions existence. We find evidence supporting the reality of illusory conjunctions (ICs), even taking into account that other conjunction errors could take place. Considering the obtained results, we conclude not only that ICs are a real phenomenon, but easier to occur in a temporal dimension than in a spatial one.

Besides, tree models fit allow to obtain more precise information about how cognitive system works in RSVP tasks. The only model that shows a good fit to participants' performance proposes two cognitive processing systems. The main system to elaborate a response - focal attention - has a slow and controlled nature (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). The system that generates floating features works in a fast and automatic way, but their generated representations would be used to give a response only when focal attention fails in reaching this objective. When fitting this model individually the results are similar to those obtained with group data, but there is a small number of participants that shows a performance that even this model fails to describe. The proportion of individuals for which this occurs was not significantly different of what could be expected in the case on which the model fit to all individuals. However, it is also plausible as explanation that there are individual differences in the cognitive strategy employed to solve the task. If this would be the case, for IC and also for any other cognitive phenomena, their results could be adequately described with different models.

In study 2 we explore this possibility simulating results of a RVSP task in presence of different sources of individual variability. We employ group and individual data treatment strategies with this simulated data before assessing two MPT models. The results obtained in homogeneity conditions are congruent with other studies (e.g. Chechile, 2009; Estes & Maddox, 2005), finding that group strategy yield more precise parameter estimations that individual one. However, when fitting models with group data without checking if i.i.d. assumption holds there is a risk of obtaining biased parameter estimates and non-representative fit results. Specifically, when there are groups of participants that follow different models there is a risk of rejecting models that adequately describe behavior of some of them. Also, models that fail to describe

behavior of some participants could be held as general explanations. If we bear in mind that group data strategy is the most widely employed in mathematical model fitting in psychology, this habit would be leading to an over-representation of some theoretical explanations in scientific literature. In other words, we could be taking for granted that some explanations are basic or general when it could be not representative of some individuals in the population. Attending to study 2 results, we propose as advisable to employ an individual fit strategy when there are cues suggesting that participants could be employing different strategies in experimental task solving.

Consequences of individual differences in model fit is not a recent research topic (see Estes, 1956), but its popularity has raised again recently. This can be seen, for example, in the MPT classical model extensions proposals, among which Klauer's (2010) and Matzke et al. (2015) hierarchical latent trait models are the most recent. Hierarchical latent trait model is designed to take account of continuous differences between participants, overcoming the limitations of the individual and group data treatment strategies. In study 3 we assessed the performance of this model in presence of different sources of individual differences. Latent trait model parameters posterior distributions are sensitive to situations on which the only source of heterogeneity is a mixture of participants that fit to different models, allowing to make inferences on the presence of those groups by discontinuities and multiple modal values. Nevertheless, in presence of subgroups with close parameter values (e.g. 0.65 and 0.70) this advantage could not be so clear. In that sense, latent trait model performance is similar to that could be obtained by fitting MPT classic models to each individual data set. Besides, the main advantage of latent trait model could disappear when within each group of participants that follow different MPT models are also continuous differences. When this is the case, identification of groups of participants that fit to different models becomes harder, or even impossible. This problem is related to another - as far as we know - non-solved one: there is no reliable method available to classify participants as following one of multiple proposed models.

The results of the studies presented here show up a problem for which there is no formal solution in psychology and that could be the source - at least partially - of some relevant controversies in different fields. One of them could be the one about the existence and degree of personal consistency in human behavior under different situational factors (Ross & Nisbett, 2011). Thus, it is possible that some individuals

were more sensitive and reactive to given situational factors than others, generating two groups of participants showing similar behavior inside each group, but qualitatively different between them when facing a similar situation. Also, it is possible that the problem raised here could be the explanation behind some of the results that gave place to the ‘replication crisis’ in psychology (Open Science Collaboration, 2015; Pashler & Wagenmakers, 2012). Given the habit of group data treatment without checking if i.i.d. assumption holds, when studying a given phenomenon there could be studies supporting its existence while others not if there are qualitatively different groups of participants in population under study.

Referencias

- Ashby, F. G., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science*, 5(3), 144-151.
- Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review*, 103, 165-192.
- Barriopedro, M. I. & Botella, J. (1998). New evidence for the zoom model using the RSVP technique. *Perception & Psychophysics*, 60(8), 1406-1414.
- Batchelder, W. & Riefer, D. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin and Review* 6, 57-86.
- Benaglia, T., Chauveau, D., Hunter, D. R., & Young, D. (2009). Mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software*, 32(6), 1-29.
- Botella, J. (2005). *Los modelos matemáticos en psicología*. Nonpublished manuscript, Facultad de Psicología, Universidad Autónoma de Madrid
- Botella, J., Barriopedro, M. I., & Suero, M. (2001). A model of the formation of illusory conjunctions in the time domain. *Journal of Experimental Psychology: Human Perception and Performance*, 27(6), 1452-1467.
- Botella, J. & Eriksen, C. W. (1991). Pattern changes in rapid serial visual presentation tasks without strategic shifts. *Bulletin of the Psychonomic Society*, 29, 105-108.
- Botella, J. & Eriksen, C. W. (1992). Filtering versus parallel processing in RSVP tasks. *Perception & Psychophysics*, 51, 334-343.
- Botella, J., Garcia, M. L., & Barriopedro, M. I. (1992). Intrusion patterns in rapid serial visual presentation tasks with two response dimensions. *Perception & Psychophysics*, 52(2), 547-552.
- Botella, J., Narváez, M., Suero, M., & Juola, J. F. (2007). Parallel processing of stimulus features during RSVP: Evidence from the second response. *Perception & Psychophysics*, 69(8), 1315-1323.

- Botella, J., Peña, D., Contreras, M. J., Shih, P., & Santacreu, J. (2009). Performance as a function of ability, resources invested, and the strategy employed. *The Journal of General Psychology*, 136(1), 41-69.
- Botella, J., Suero, M., & Durán, J. I. (2017). On the reality of illusory conjunctions. *Journal of General Psychology*, 1-19.
- Castiello, U. & Umiltà, C. (1990). Size of the attentional focus and efficiency of processing. *Acta Psychologica*, 73(3), 195–209.
- Cavagnaro, D. R., Myung, J. I., & Pitt, M. A. (2013). Mathematical modeling. In Little, T. D. (ed.) *The Oxford handbook of quantitative methods*. New York: Oxford University Press.
- Chechile, R. A. (2009). Pooling data versus averaging model fits for some prototypical multinomial processing tree models. *Journal of Mathematical Psychology*, 53, 562-573.
- Cohen, A. L., Sanborn, A. N., & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review*, 15(4), 692-712.
- Donk, M. (1999). Illusory conjunctions are an illusion: The effects of target-nontarget similarity on conjunction and feature errors. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1207-1233.
- Donk, M. (2001). Illusory conjunctions die hard: A reply to Prinzmetal, Diedrichsen, and Ivry (2001). *Journal of Experimental Psychology: Human Perception & Performance* 27, 543-546.
- Durán, J. I., Suero, M., & Botella, J. (2015). *Comparación de modelos multinomiales de procesamiento en árbol con estrategias de ajuste individual y grupal: El caso de las conjunciones ilusorias*. (Dissertation/master's thesis). Universidad Autónoma de Madrid, Facultad de Psicología, España.
- Erdfelder, E., Auer, T. S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial Processing Tree Models; A Review of the Literature. *Zeitschrift für Psychologie / Journal of Psychology*, 217, 108-124.
- Eriksen, C. & St James, J. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, 40(4), 225-40. doi: 10.3758/BF03211502

- Eriksen, C. W. & Yeh, Y.-Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception & Performance*, 11, 583-597.
- Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, 53(2), 134-140.
- Estes, W. K. & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review*, 12(3), 403-408.
- Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology*, 59, 255–278.
- Gelman, A. & Hill J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.
- Gluck, J. & Fitting, S. (2003). Spatial strategy selection: Interesting incremental information. *International Journal of Testing*, 3, 293-308.
- Gluck, J. Machat, R. Jirasko, M. & Rollett, B. (2001). Training-related changes in solution strategy in a spatial test: An application of item response models. *Learning & Individual Differences*, 13, 1-22.
- Green, D. M. & Swets, J. (1988). *Signal detection theory and psychophysics*. Nueva York: Wiley.
- Heathcote, D., Walker, P., & Hitch, G. J. (1994). Feature independence and the recovery of feature conjunctions. *The Journal of General Psychology*, 121(3), 253-266.
- Houpt, J. W., Yang, C-T., & Townsend, J. T. (2016). Modeling individual differences in perceptual decision making. *Frontiers in Psychology*, 7.
- Hu, X. & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, 59(1), 21-47.
- Ivry, R. B., & Prinzmetal, W. (1991). Effect of feature similarity on illusory conjunctions. *Perception & Psychophysics*, 49, 105-116.
- Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate Discrete Distributions*, 3rd ed. New Jersey: John Wiley & Sons.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Strauss and Giroux.

- Kawahara, J. I., & Miyatani, M. (2001). The effect of informative and uninformative cueing of attention on feature integration. *The Journal of general psychology*, 128(1), 57-75.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and psychological measurement*, 56(5), 746-759.
- Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent class approach. *Psychometrika*, 71(1), 7-31.
- Klauer, K. C. (2010). Hierarchical multinomial processing tree models: a latent-trait approach. *Psychometrika*, 75, 70-98.
- Lawrence, D. H. (1971). Two studies of visual search for word targets with controlled rates of presentation. *Perception & Psychophysics*, 10, 85-89.
- LaBerge, D., Carlson, R. L., Williams, J. K., & Bunney, B.G. (1997). Shifting attention in space: Tests of moving spotlight models vs an activity-distribution model. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 1380-1392.
- Lee, M. D. & Webb, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin & Review*, 12(4), 605-621.
- Lemaire, P. & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, 124, 83-97.
- Lunn, D., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10, 325-337.
- Luce, R. D. (1959). *Individual Choice Behavior*. Nueva York: Wiley.
- Marewski, J. N. & Olsson, H. (2009). Beyond the null ritual. Formal modeling of psychological processes. *Zeitschrift für Psychologie/Journal of Psychology*, 217(1), 49-60.
- Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, 80(1), 205-235.

- McLachlam, G. & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Myung, J. I., Cavagnaro, D. R., & Pitt, M. A. (2017). Model evaluation and selection. In Batchelder, W. H., Colonius, H., Dzhafarov, E., and Myung, J. I. (eds) *New Handbook of Mathematical Psychology, Volume I: Measurement and Methodology*. Cambridge, UK. Cambridge University Press.
- Myung, I. J., Pitt, M. A., & Kim, W. (2005). Model evaluation, testing and selection. In Lambert, K. & Goldstone, R. (eds.) *Handbook of cognition*. London: Sage Publication
- Navarro, D. J., Griffiths, T. L., Steyvers, M., & Lee, M. D. (2006). Modeling individual differences using dirichlet processes. *Journal of Mathematical Psychology*, 50, 101-122.
- Niu, X., Li, P., & Zhang, P. (2011). Testing homogeneity in a multivariate mixture model. *The Canadian Journal of Statistics*, 39(2), 218-238.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251).
- Pashler, H. & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science a crisis of confidence? *Perspectives on Psychological Science*, 7(6), 528-530.
- Peña, D., Contreras, M. J., Shih, P. C., & Santacreu, J. (2008). Solution strategies as possible explanations of individual and sex differences in a dynamic spatial task. *Acta Psychologica*, 128, 1-14.
- Ponsoda, V. (1986). *Iniciación a la Psicología Matemática*. Madrid: Servicio de Publicaciones de la UAM.
- Prinzmetal, W., Diedrichsen, J., & Ivry, R. (2001). Illusory conjunctions are alive and well: A reply to Donk (1999). *Journal of Experimental Psychology: Human Perception & Performance*, 27, 538-541.
- Quinlan, P. (2003). Visual Feature Integration Theory: Past, Present, and Future. *Psychological Bulletin*, 129(5), 643-673.

- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, 19, 90-137.
- Reder, L. M. & Schunn, C. D. (1999). Bringing together the psychometric and strategyworlds: Predicting adaptivity in a dynamic task. In Gopher, D. & Koriat A. (eds.) *Cognitive regulation of performance: Interactions of theory and applications. Attention and performance, XVII*. Cambridge, MA: MIT Press.
- Riefer, W. & Batchelder, W. H. (1988). Multinomial modelling and the measurement of cognitive processes. *Psychological Review*, 95(3), 318-339.
- Riefer, W., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, D. (2002). Cognitive Psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, 14, 184-201.
- Ross, L. & Nisbett, R. E. (2011). The search for personal consistency. In Ross, L. & Nisbett, R. E. (2011). *The person and the situation. Perspectives of social psychology*. London: Printer and Martin.
- Rouder, J. N. & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Bulletin & Review*, 12(4), 573-604
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime user's guide. Pittsburgh, PA: Psychology Software Tools, Inc.
- Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing. I. Detection, search, and attention. *Psychological Review*, 84(1), 1-66.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing. II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84(2), 127-190.
- Siegler, R. S. & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General*, 126, 71-92.

- Singmann, H. & Kellen, D. (2013). MPTinR: Analysis of Multinomial Processing Tree models in R. *Behavior Research Methods*, 45, 560-575.
- Smith J. B. & Batchelder, W. H. (2008). Assessing Individual Differences in Categorical Data. *Psychonomic Bulletin and Review*, 15(4), 713-731.
- Smith J. B. & Batchelder, W. H. (2010). Beta-MPT: Multinomial Processing Tree Models for Addressing Individual Differences. *Journal of Mathematical Psychology*, 54, 167-183.
- Speelman, C. P. & Muller-Townsend, K. L. (2015). Attaining automaticity in the visual numerosity task is not automatic. *Frontiers in Psychology*, 6.
- Stevens, S. (1975). *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects*. John Wiley & Sons, New York.
- Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software*, 12(3), 1-16.
- Treisman, A. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97-136.
- Treisman, A. & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107-141.